Electronographic polarimetry of reflection nebulae

R. F. Warren-Smith

ELECTRONOGRAPHIC POLARIMETRY OF REFLECTION NEBULAE

by

R. F. WARREN-SMITH, B.A.

A THESIS SUBMITTED TO THE UNIVERSITY OF DURHAM FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

August, 1979.

ABSTRACT

In the first part of this thesis an imaging polarimeter developed at the University of Durham for use in the optical waveband is described, and details are presented of a new computer data reduction system which allows the production of polarisation maps automatically from polarimeter images recorded using a McMullan electronographic camera. New methods are employed to achieve higher accuracy and reliability than previously possible, particularly at low brightness levels. A method of checking the smoothness and mutual consistency of images is described, and its use in eliminating image defects is illustrated. This technique is also used to investigate the sources of error in electronography and ways are suggested in which these errors can be reduced.

In the second part of the thesis, new polarimetric observations of the reflection nebula NGC1999 are presented and discussed in the light of previous colour and brightness observations. Theoretical arguments are used to relate important nebular parameters to observational features in reflection nebulae, and these ideas are applied to produce a numerical model of NGC1999. Complete multiple scattering calculations are performed and the model is shown to be consistent with all available observations of the nebula. Finally, observations of other reflection nebulae are discussed and the experience gained with NGC1999 is used to outline deficiencies in current models of the interstellar extinction law, and it is suggested that a broad spectrum of particle sizes in the interstellar medium may be indicated. The importance of imaging polarimetry in further investigation of this problem is pointed out.

PREFACE

The contents of this thesis describe work undertaken by the author during a three year period between 1976 and 1979 while he was a research student at the University of Durham, under the supervision of Dr. S. M. Scarrott.

The work of the Astronomy group, in which the author was involved, concerns the study of optical polarisation in extended astronomical objects and the measurement of this polarisation using an instrument developed jointly at Durham and the Royal Greenwich Observatory. During the author's time in Durham he was frequently involved in the use of this polarimeter at observing sites abroad.

Although the author had little to do with the development of the present polarimeter, he was principally concerned in the development of a computer data-reduction technique for handling the recorded polarimeter images. The description of this technique is necessarily involved, including much numerical detail, and adds considerably to the bulk of this volume. The inclusion of this detail was thought necessary, however, partly because many of the new techniques employed are relevant to the reduction of similar data in other contexts, and partly because a detailed description of the processes involved is clearly needed as documentation for future users of the computer programmes written for this work. It is hoped that the necessary balance between detail and brevity has been struck. Much of the formidable task of computer programming involved in the reduction technique was

carried out in close collaboration with Dr. W. S. Pallister of the University of Durham.

The later chapters of the thesis are concerned with work undertaken by the author on the interpretation of polarisation measurements, made with the Durham polarimeter, on reflection nebulae. Data obtained by the author for the nebula NGC1999 at the Anglo-Australian Telescope in January 1978 forms a basis for much of this work.

CONTENTS

PART 1

ELECTRONOGRAPHIC POLARIMETRY

CHAPTER 1

THE DESCRIPTION OF POLARISED LIGHT

1.1 The Nature of Polarisation

The essential feature of light, which distinguishes it from scalar waves such as sound, and enables it to show polarisation phenomena, is that the disturbances of the electric and magnetic fields which constitute light are vector quantities. Consequently, we need more parameters to describe a beam of light than we do to describe a sound wave. Exactly which of the several possible sets of parameters we chose to use depends on the application, and to see the equivalence of these parameters requires an understanding of the nature of polarisation in light.

As with sound waves, we may consider a general electro-magnetic disturbance as a superposition of many waves travelling in different directions with a range of frequencies and associated amplitudes and phases. Each of these elementary waves must satisfy the wave equation, derivable from Maxwell's equations (see e.g. Reitz and Milford, 1973), which shows that the electric and magnetic vectors in the wave are mutually perpendicular, and perpendicular to the direction of wave propagation (Figure 1.1). In the case of light, therefore, there are an infinite number of these elementary waves, differing only in the direction of the electric and magnetic vectors within the plane normal to the propagation direction.

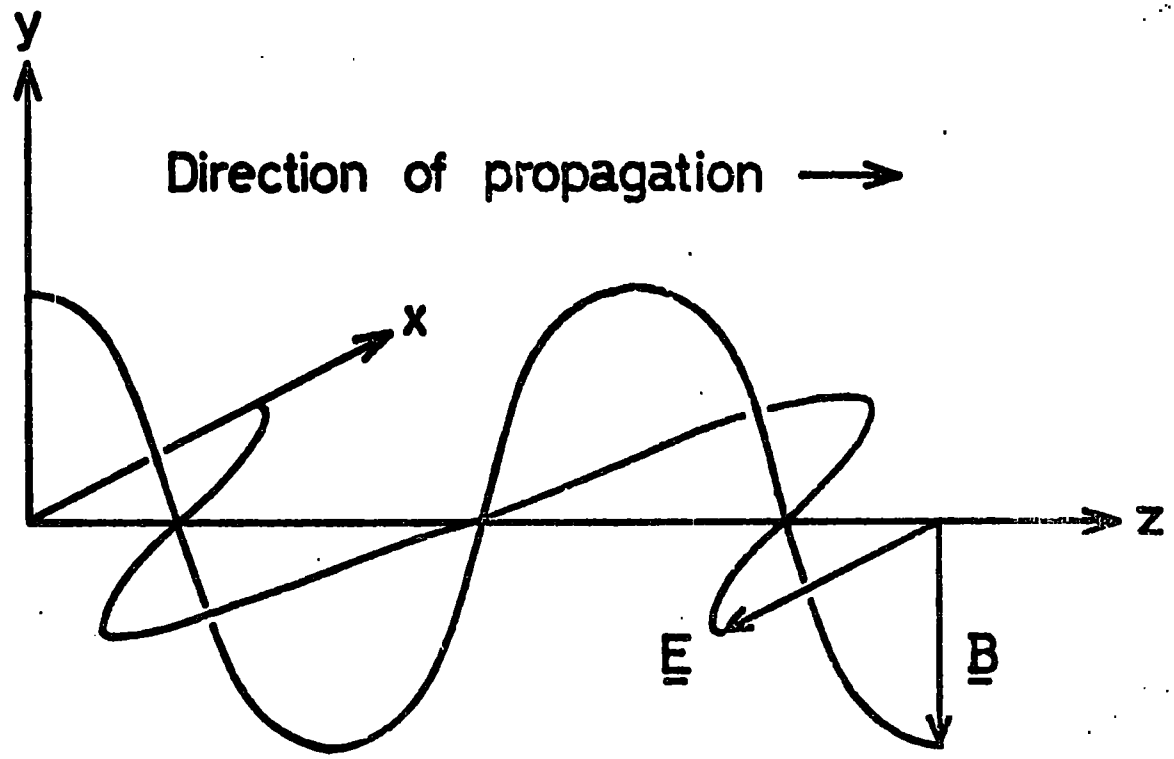
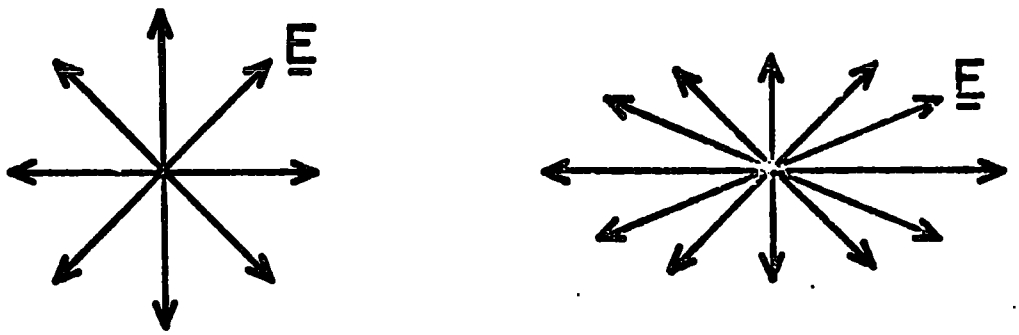


Fig. 1.1 A simple plane-polarised wave



a) Unpolarised incoherent light.

b) Horizontally plane polarised light.

Fig. 1.2 Partial plane polarisation

Natural, incoherent light consists of a superposition of many of these elementary waves, with their electric and magnetic vectors in differing directions. If the electric vectors are distributed randomly, and show no preferred direction, the light is said to be unpolarised. If there is a preferred direction the light is said to be plane polarised in this direction. The usual convention is to use the electric vector to refer to the direction of polarisation, although the magnetic vector is sometimes used (Figure 1.2).

So far we have considered only elementary waves with the electric vector in a fixed direction. In fact any linear sum of these elementary waves may be used instead, and we can produce all the possible combinations by considering the resultant of two waves plane polarised perpendicularly and with a phase difference between them. Using the usual complex exponential notation for wave propagation, such a wave, propagating in the $+z$ direction, has an electric vector $\underline{E}$ of the form:

$$\underline{E} = \mathrm{Re}\left[\underline{p}\, E_x \exp(i\{\omega t - kz\}) + \underline{q}\, E_y \exp(i\{\omega t - kz\} + \delta)\right] \quad (1.1)$$

where $\underline{p}$, $\underline{q}$ are unit vectors in the x, y directions and E_x, E_y are complex amplitudes for the waves polarised in these directions (Figure 1.3). Since the absolute phase is unimportant, we in fact need three quantities to fully describe an elementary light wave: two real amplitudes and a phase difference d, compared with only one amplitude for a sound wave.

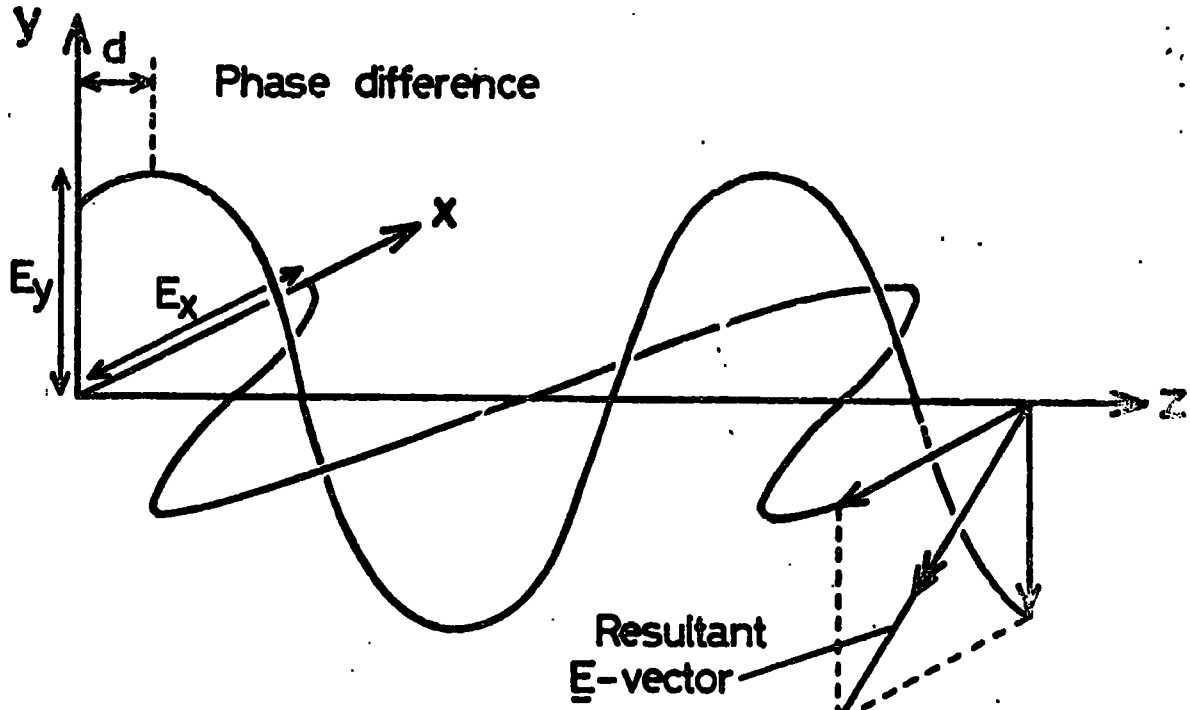


Fig. 1.3 General polarisation of a simple wave

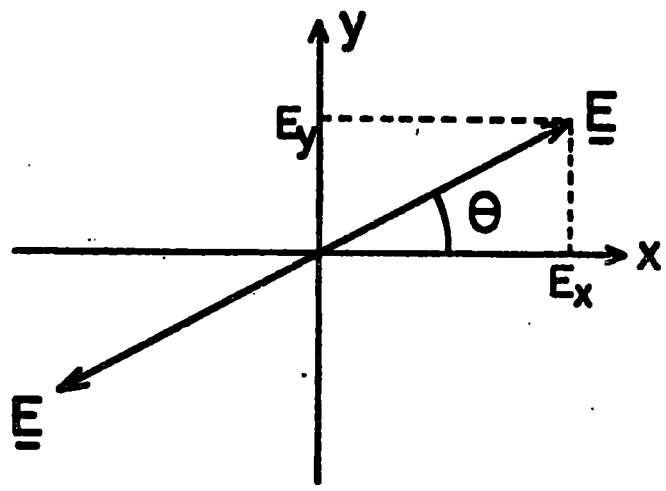


Fig. 1.4 Plane polarisation

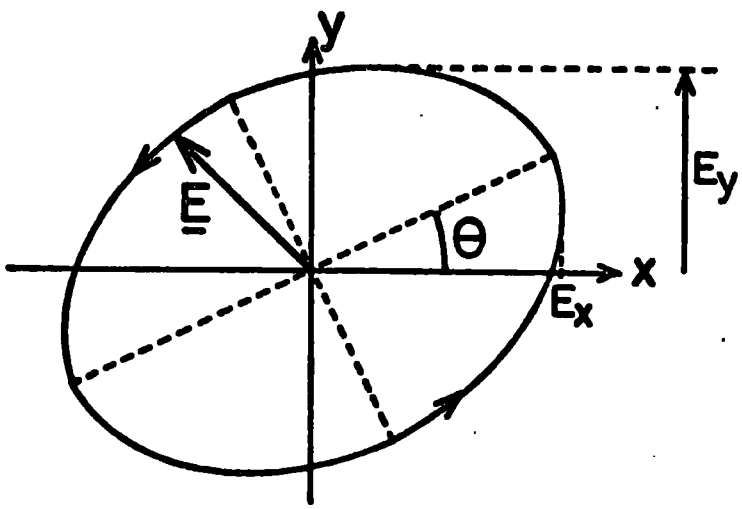


Fig. 1.5 Elliptical polarisation

If d = o, the electric vector is still confined to one direction (Figure 1.4). If d ≠ o, the wave is, in general, elliptically polarised and the electric vector rotates to trace out an ellipse with time (Figure 1.5). By convention this polarisation is called right handed if the electric vector rotates clockwise with time when viewed along the direction of propagation. The electric vector then describes a left handed helix in space as a function of z (Figure 1.7). In general this helix has an elliptical cross section, but if $d = \pm\lambda/4$, where λ is the wave-length, and $|E_x| = |E_y|$, the ellipse becomes a circle and the polarisation is described as circular (Figure 1.6).

1.2 The Stokes Parameters

Perfectly coherent waves such as we have considered so far must necessarily have one of these pure forms of polarisation. However, while a telescope can resolve light into components in different directions, it cannot resolve perfectly into frequency components, so the light which we measure will contain a number of independent components, i.e. it will be incoherent. We must therefore consider the effect of having a large number of independent polarised waves with different frequencies and random phases. Because of the random nature of this incoherent light we consider time averages, denoted by $\langle\ \rangle$, and following van de Hulst (1957) we

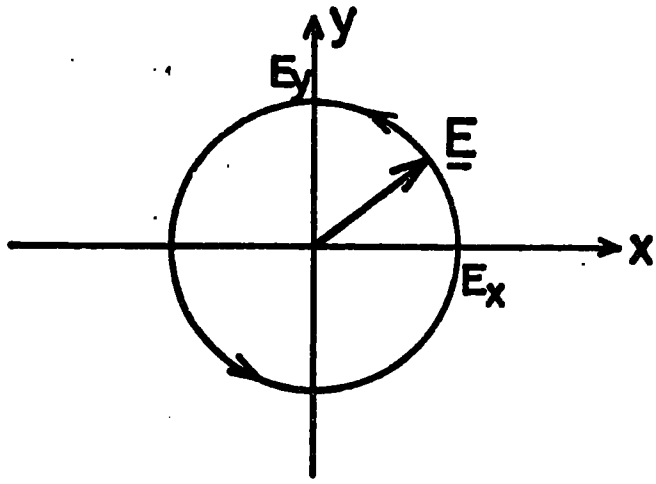


Fig. 1.6 Circular polarisation

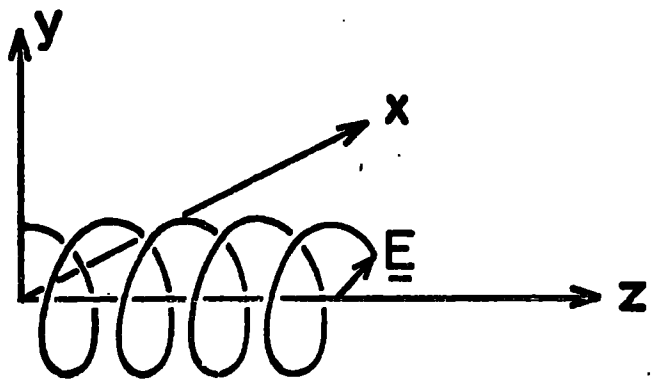


Fig. 1.7 The electric vector in space

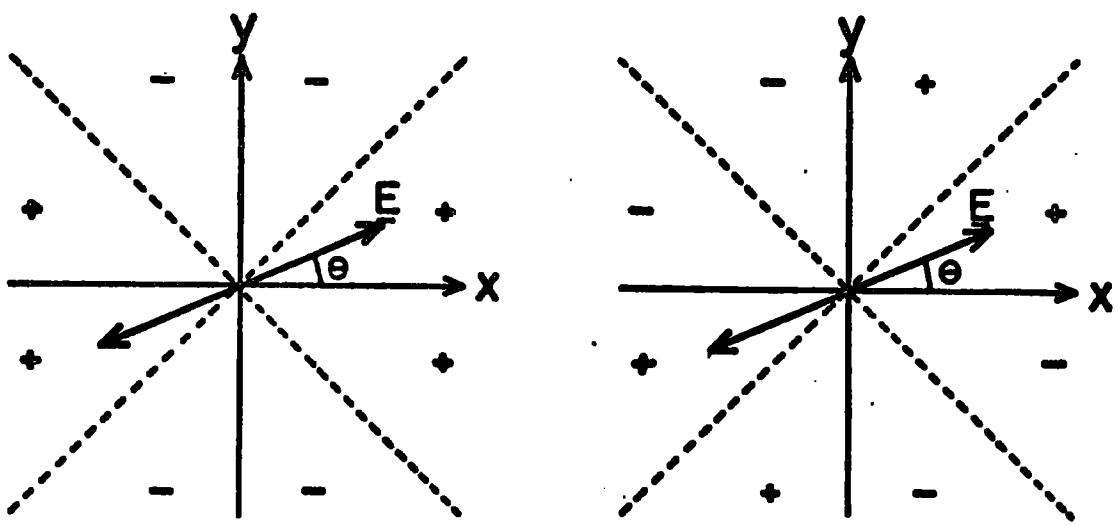


Fig. 1.8 The signs of the Stokes parameters

define the Stokes parameters of a beam of light propagating in the $+y$ direction as:

$$I = \langle E_x E_x^* + E_y E_y^* \rangle \qquad 1.2$$

$$Q = \langle E_x E_x^* - E_y E_y^* \rangle \qquad 1.3$$

$$U = \langle E_x E_y^* + E_y E_x^* \rangle \qquad 1.4$$

$$V = i \langle E_x E_y^* - E_y E_x^* \rangle \qquad 1.5$$

These parameters are sometimes considered as components of a Stokes vector:

$$\underline{S} = (I, Q, U, V) \qquad 1.6$$

They are real quantities with the dimensions of intensity, which, for an elementary wave, satisfy:

$$I^2 = Q^2 + U^2 + V^2 \qquad 1.7$$

Thus, for an elementary wave, only three of the Stokes parameters are required to fully specify it.

I is the total intensity of the light beam, while Q and U describe the state of linear polarisation, Q representing the preference of the electric vector for the $\theta = 0$ or $\theta = \pi/2$ directions (Figure 1.8) and U representing its preference for the $\theta = \pm \pi/4$ directions. V describes the preference for right or left handed circular polarisation. Evaluation of the Stokes parameters for an elementary wave (Equation 1.1) gives:

$$I = |E_x|^2 + |E_y|^2 \qquad 1.8$$

$$Q = |E_x|^2 - |E_y|^2 \qquad 1.9$$

$$U = 2\,|E_x|.\,|E_y| \cos kd \qquad 1.10$$

$$V = 2\,|E_x|.\,|E_y| \sin kd \qquad 1.11$$

Furthermore, it can be shown that the orientation of the major axis of the ellipse (Figure 1.5) is given by:

$$\tan 2\theta = U/Q \qquad 1.12$$

In the case where V = o, the wave is plane polarised and this formula gives the orientation of the plane polarisation (Figure 1.4). If a new set of x,y axes is chosen, by rotating the reference frame about the z axis, it can be shown that the quantities

$$I, \quad \sqrt{Q^2 + U^2}, \quad V$$

are unchanged. For a fuller discussion see van de Hulst (1957).

The prime usefulness of the Stokes parameters can be appreciated by considering the superposition of a large number of incoherent waves $\underline{E}_i$ with Stokes vectors $\underline{S}_i$. If the waves are independent such that:

$$\langle E_{xi} E_{xj}^* \rangle = \langle E_{xi} E_{yj}^* \rangle = \langle E_{yi} E_{yj}^* \rangle = o \text{ for } i \neq j \qquad 1.13$$

then the Stokes vector representing the composite wave $\underline{E}_R$, given by:

$$\underline{E}_R = \sum_i \underline{E}_i \qquad 1.14$$

is the sum of the individual Stokes vectors:

$$\underline{S}_R = \sum_i \underline{S}_i \qquad 1.15$$

That is, the Stokes parameters for independent beams of light are additive.

1.3 The Degree of Polarisation

In general, for a beam of light composed of many elementary waves, equation 1.7 is no longer valid. Instead we have:

$$I^2 \geqslant Q^2 + U^2 + V^2 \qquad 1.16$$

the equality only being true if all the elementary waves have similar polarisations. Since I describes the total intensity and Q, U, V describe the polarisation, we can define a polarised intensity:

$$Ip = \sqrt{(Q^2 + U^2 + V^2)} \qquad 1.17$$

which leaves an unpolarised component:

$$I_u = I - I_p \qquad 1.18$$

We can also define the degree of polarisation:

$$P = (Q^2 + U^2 + V^2)^{\frac{1}{2}} / I = I_p / I \qquad 1.19$$

This quantity is often expressed as a percentage and takes values between 0% for an unpolarised beam and 100% for a fully polarised beam. As can be seen from equation 1.7, pure elementary waves are always 100% polarised, and some form of incoherence is necessary to reduce this figure.

1.4 Optical Equivalence

The Stokes parameters can be used to illustrate the important principle of optical equivalence. Consider any linear optical instrument with an elementary input wave with x, y components A_x , A_y , and an output wave

with components B_x , B_y . Since the instrument is linear:

$$B_x = a_1 A_x + a_2 A_y \qquad 1.20$$

$$B_y = a_3 A_x + a_4 A_y \qquad 1.21$$

where a_1 to a_4 are a set of coefficiants dependent on the instrument. If we now calculate the Stokes parameters of the output beam, we find that they are related to those of the input beam by a linear transformation dependent on a_1 to a_4. Since measurements of polarisation are limited to the use of a linear optical instrument followed by a measurement of the total intensity I of the output beam, it follows that we cannot distinguish between two input beams which have the same Stokes parameters. The Stokes parameters are therefore all that one can determine about the polarisation of a beam of light, even if it can be composed of elementary waves in an infinite number of ways. Measurements of polarisation are therefore directed at determining the Stokes parameters (or an equivalent set of four parameters), and in the case of the Durham University polarimeter, those relating to plane polarisation of the light, namely I, Q, U, the measurements being insensitive to the value of V.

1.5 The Quantum Description of Polarisation

Since it is sometimes necessary to refer to individual photons, brief mention should be made of the quantum

description of polarisation in light.

Photons may be thought of as a special case of spin 1 particles, where the spin up and spin down ($m = \pm 1$) states are possible, but the usual $m = 0$ state is excluded. The polarisation state vector Ψ of an individual photon is given, in accordance with the laws of quantum mechanics, as a linear sum of 2 orthogonal vectors B_1, B_2 :

$$\Psi = c_1 B_1 + c_2 B_2 \qquad 1.22$$

where c_1, c_2 are complex amplitudes, and the basis vectors may be chosen for convenience, e.g. spin up and spin down, orthogonal linear polarisation states, etc. Any measurement of polarisation which forces this photon into another polarisation state ϕ, also expressable as a linear sum of B_1, B_2 (e.g. resolution into a linear polarisation state using a sheet of polaroid), will give a result (transmission or absorption) which is unpredictable in advance, but which has a probability given by the usual quantum mechanical formula $\langle \Psi | \phi \rangle$ for transmission. This resolution into polarisation states may be seen as an analogue of the Stern-Gerlach experiment for spin ½ particles. For a fuller discussion of the quantum description of polarisation see e.g. Feynman (1965).

CHAPTER 2

THE NEBULA POLARIMETER AND ELECTRONOGRAPHIC CAMERA

2.1 Introduction

The Durham University nebula polarimeter was designed to measure the degree of linear polarisation in the light from extended astronomical objects and to display the spatial structure of this polarisation in the form of a polarisation map.

The principles of the instrument have changed little since the prototype was built by Scarrott, Bingham and Axon (Axon 1977), based on a design by Ohman (1939). There have, however, been several practical realisations of the instrument, differing in mechanical construction and the type of optics employed. The original design philosophy has been adequately described by several authors (Axon 1977 , Pallister 1976), so the description presented here will concentrate on the most recent instrument, pointing out any minor improvements on previous versions.

2.2 Polarimeter Principles

The principle of any polarimeter is to resolve the incoming light into known polarisation states and to record the resulting intensity. At least 3 such measurements on different linear polarisation states must be made to fully determine the degree of linear

polarisation of light. This is usually achieved by rotating the instrument, or some component of it, and measuring at several different orientations. Practical considerations, however, dictate that more than 3 measurements must be made, and it is here that different polarimeter design philosophies emerge.

The results produced will be erroneous if the transparency of the atmosphere changes with time, unless corrections are made for this. To make these corrections it is necessary to record two orthogonal polarisation states simultaneously, increasing the total number of measurements to 4, and this is the principle of dual beam polarimeters. Such polarimeters have the additional advantage that all the incident light is used, resulting in higher efficiency; the Durham polarimeter is of this type.

An alternative to the dual beam system is to switch rapidly between orthogonal polarisation states on a time scale much shorter than the variations in atmospheric transparency. Such A.C. polarimeters eliminate the atmospheric effects but are less efficient than dual beam types since 50% of the light is unused.

Further problems arise if the light from the night sky (which is superimposed on the light from the object and must be subtracted from the total signal) is not of a constant brightness and polarisation. This is

usually the case, except under the very best observing conditions, and it is then essential to monitor the sky at the same time as the object. Photo-electric polarimeters which measure only one region at a time cannot achieve this, and systematic sequences of observations of object and sky must be made to minimise and check for errors from this cause. Since this involves moving the telescope during the observations, it introduces the possibility that the instrument may not return to the same part of the object each time, introducing further errors. The Durham polarimeter, being an imaging device, has a sufficiently wide field that both object and sky signals can be accommodated and measured simultaneously. Furthermore, the separate images which constitute the measurements can be computer-aligned at the data-reduction stage, thus ensuring that identical parts of the object are measured each time.

The final problem which affects the measurements is the sensitivity of the detection system to different states of polarisation. If the response is not equal for orthogonally polarised beams of light, due to differential losses in the optics, or because the detector is polarisation sensitive, then further measurements are needed to determine this and make corrections for it. In the case of the Durham polarimeter this has the effect of doubling the number of measurements required to 8. Since they are made in pairs, this means that 4 separate

integration times are required to determine the polarisation. The problems associated with these corrections and with the spatial calibration of the detector will be dealt with in the description of the reduction technique (Chapter 5).

2.3 The Durham Polarimeter

2.3.1 The Polarimeter Optics

The optical system of the Durham polarimeter is shown in figure 2.1. It is designed to operate at an f 13.5 telescope focus, this being necessary to ensure that all the light from the telescope passes through the polarimeter optics.

The focal plane of the telescope is co-incident with the grid assembly of the polarimeter, shown in more detail in figure 2.2. This assembly defines the field of the polarimeter as 40 mm. in the focal plane, and consists of a series of alternate opaque bars and gaps, of equal width, which admit 50% of the telescope image to the polarimeter. This then allows space for the production of two images by double refraction in the quartz prism (Figure 2.3).

The grid assembly is machined from perspex and painted black to prevent the possibility of polarisation of the light by metallic edges, and the edges of the bars are bevelled at 45^{o} to prevent a knife-edge in the

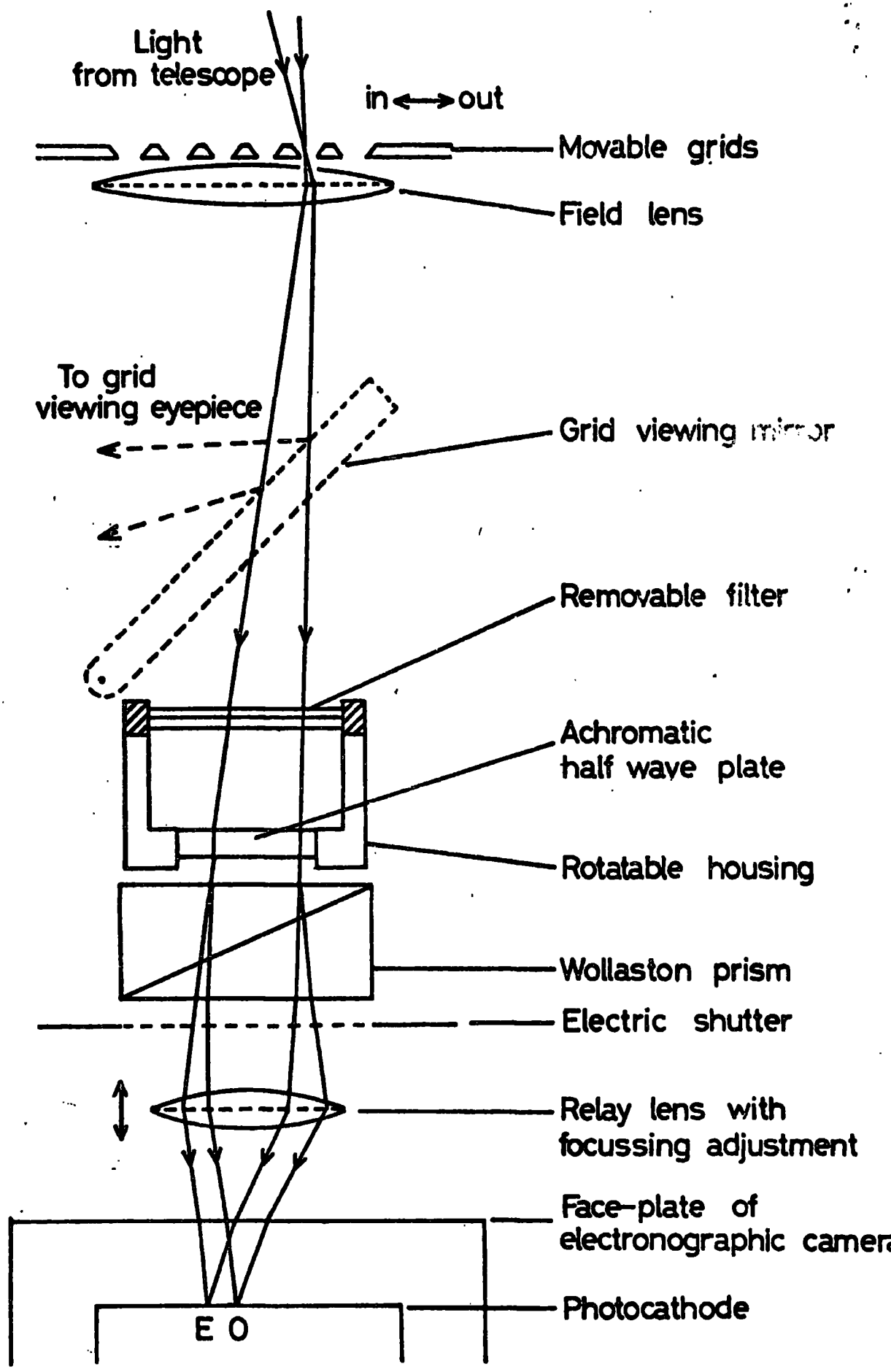


Fig. 2.1 The polarimeter optics

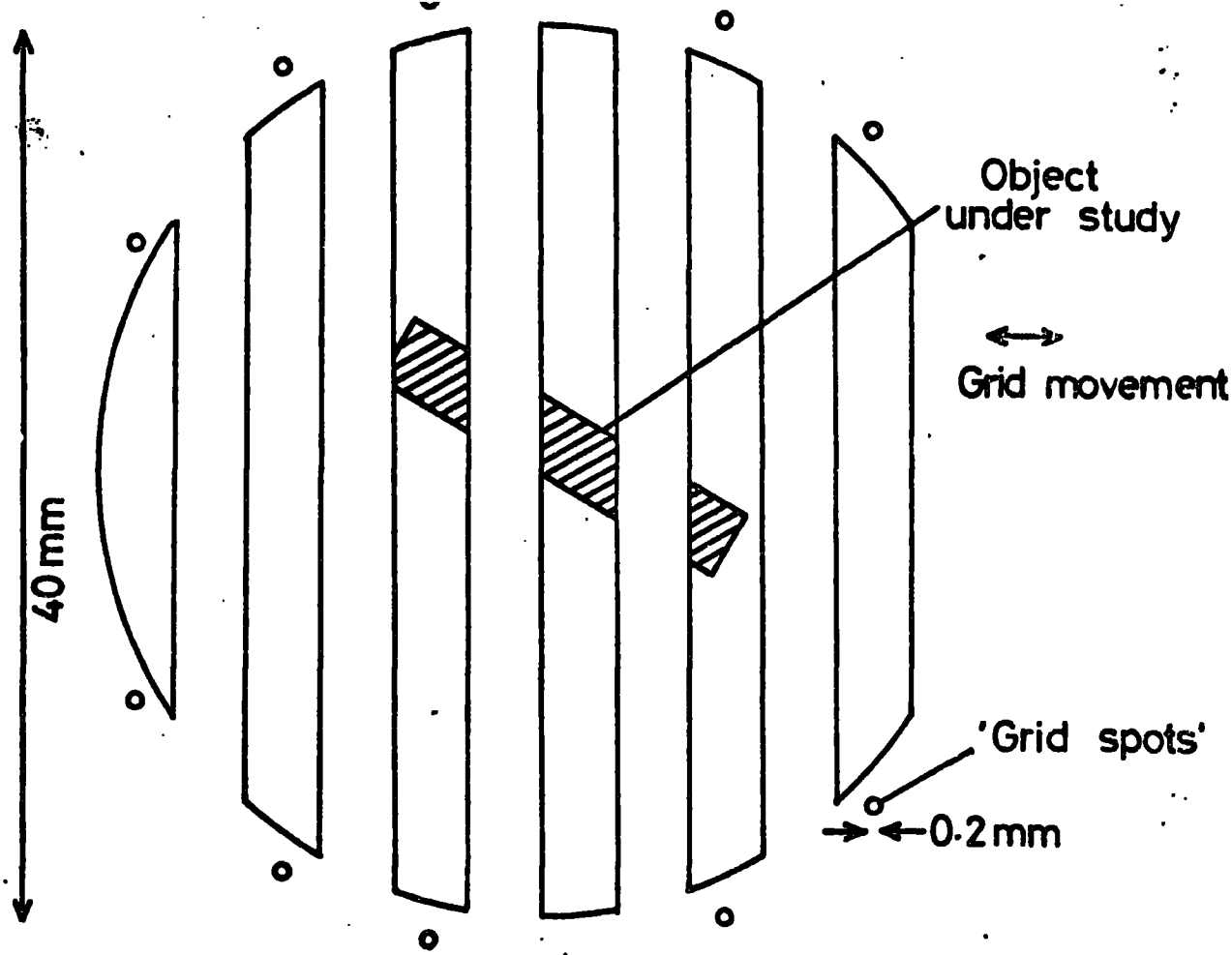


Fig. 2·2 The grid assembly

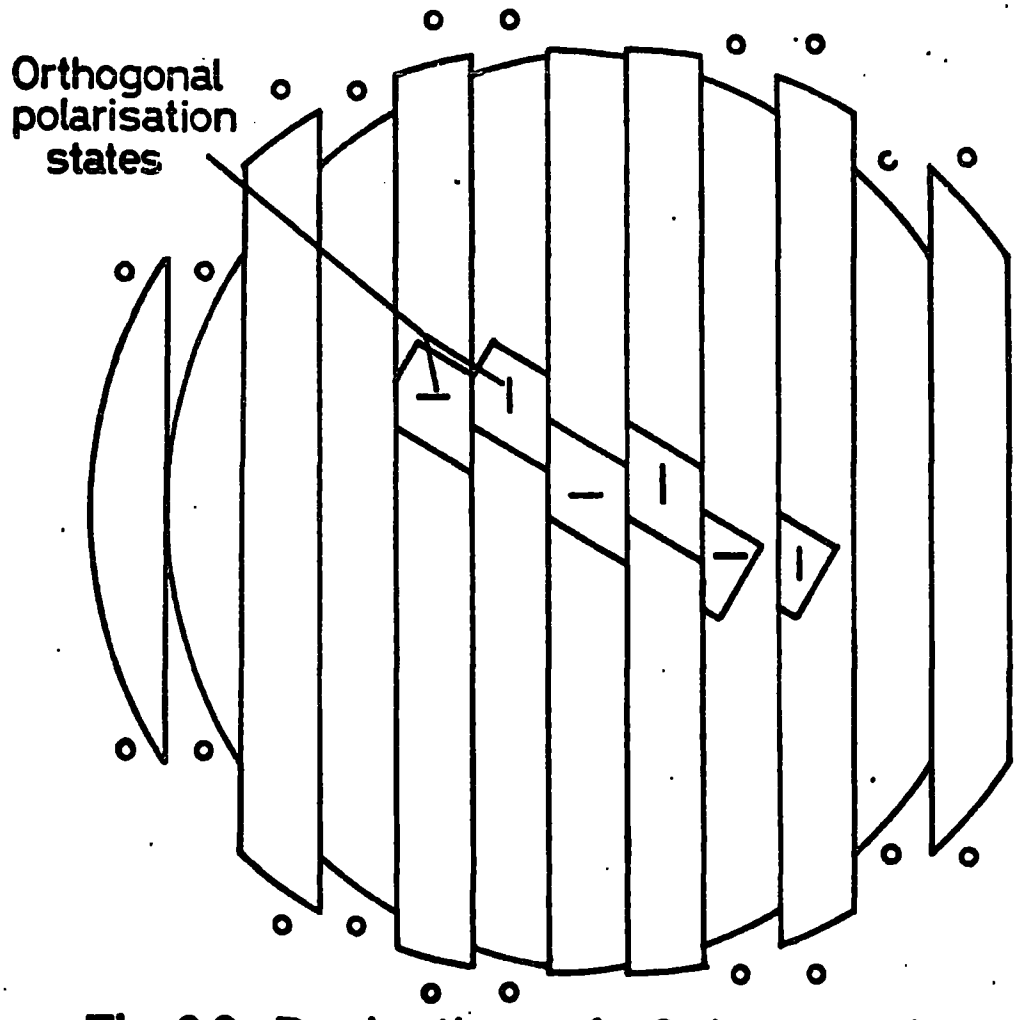


Fig. 2·3 Production of 2 images by double refraction

focal plane, thus preventing unwanted diffraction effects. The whole assembly is mounted on a slide between two stops, allowing the other 50% of the image to be viewed by moving the grid sideways by the width of one bar. Previous versions of this instrument (Axon, 1977; Pallister, 1976) have included fine wires in the grid plane for alignment and focussing, but these have now been dispensed with in favour of a series of small (0.2 mm. diameter) holes, called 'grid spots', around the field which serve both for focussing and as reference marks for image alignment during data-reduction. This leaves the whole field clear for polarimetry.

Immediately behind the grid assembly is a field lens which serves to reduce the width of the light beam sufficiently for it to pass through the remainder of the optical components. The lens is a 55mm. diameter achromatic doublet with a focal length of 180mm. It is mounted on the grid assembly with a screw fit for easy removal and cleaning, since it lies close to a focal plane, and must be kept free of dust.

A silvered perspex mirror can be inserted at 45^{o} into the optical path, behind the field lens, to deflect the light to an eyepiece in the side of the polarimeter, through which the grid assembly and telescope image can be viewed for acquisition and focussing. When the mirror is removed from the beam, the light passes on through a rotatable assembly holding a filter, which defines the

optical bandpass of the instrument, and a half-wave plate.

The effect of a half wave plate is to introduce a phase difference of 180^{o} between the components of light polarised parallel and perpendicular to its optic axis. The result of this is illustrated in figure 2.4 - the plane of polarisation of linearly polarised light is rotated by an angle 2θ, where θ is the angle which the incident polarisation makes with the optic axis. Thus by rotating a half wave plate in the beam we simulate the effect of rotating the polarisation of the incoming light, or equivalently of rotating the entire instrument.

A simple half wave plate consists of an optically active crystal, cut parallel to its optic axis, with refractive indices n_o and n_e for the ordinary and extra-ordinary rays. After travelling through a length of crystal t, they have an optical path difference of $t(n_o - n_e)$, which, for a half wave plate, must have the value $(m + \frac{1}{2})\lambda$ where m is an integer. Obviously this is only true for discrete wavelengths if n_o and n_e are constants. In order to use light over the entire optical waveband an achromatic half wave plate is required. This consists of two thick crystals bonded together. Individually they each have a large value for m, but they have $(n_o - n_e)$ of opposite sign so that their effects nearly cancel, having the overall effect of a half wave plate with a small value for m. n_o and n_e are slowly varying functions of wavelength for most crystals, but by having a large thickness of crystal this variation can have sufficient effect to

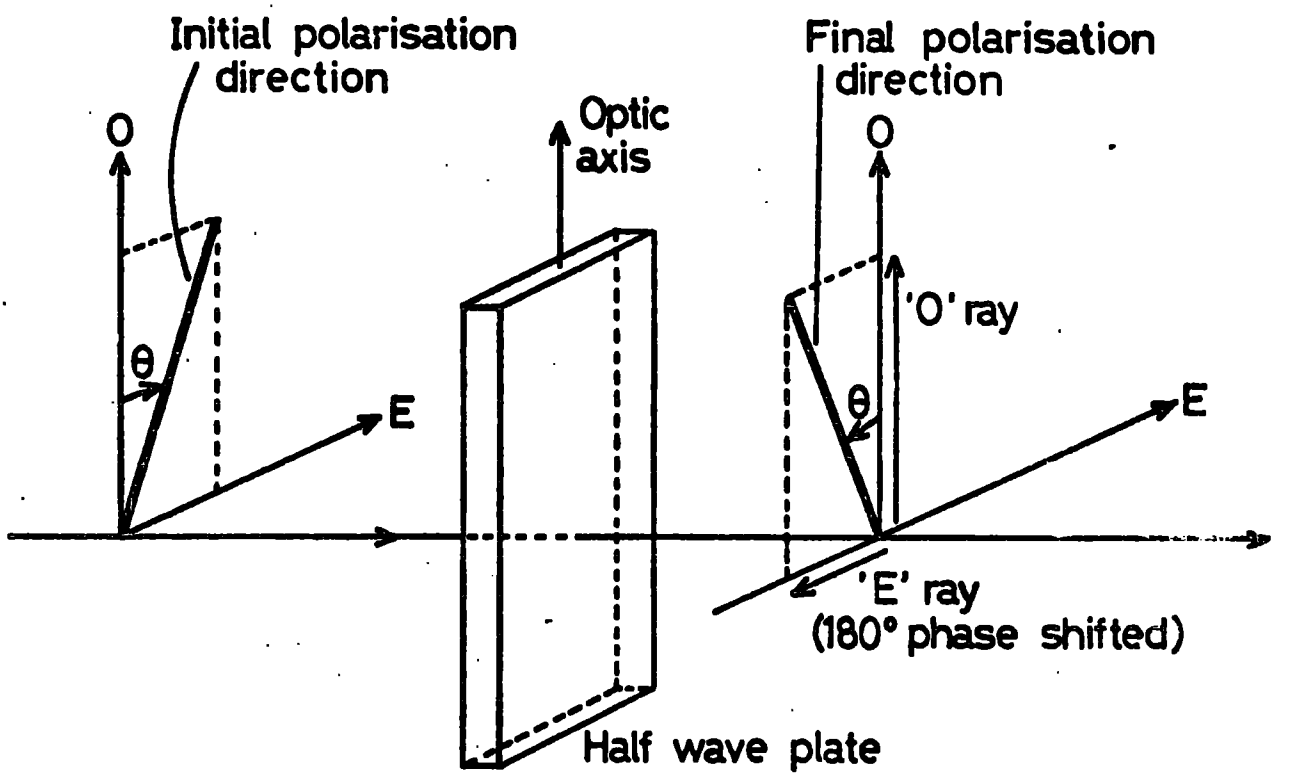


Fig. 2.4 The action of the half-wave plate

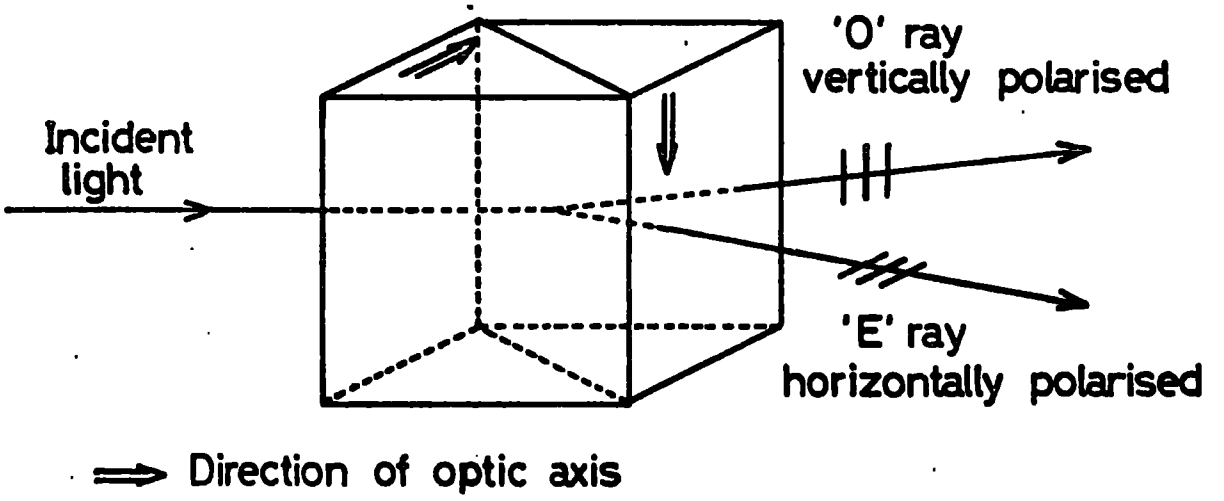


Fig. 2.5 The action of the Wollaston prism

compensate for wavelength changes producing a half wave plate effective over a range of wavelengths.

Two achromatic half wave plates are currently in use with the polarimeter: a "visual" plate, designed to operate in the wavelength range 0.48μm to 0.68μm and a "blue" plate operating between 0.39μm and 0.49μm. Both plates are of magnesium fluoride and quartz 20mm in diameter, with an acceptance angle of 8^{o}.

The assembly holding the filter and half wave plate is rotatable in fixed steps of $22\frac{1}{2}^{o}$ using a click-stop mechanism operated by gears from outside the polarimeter; thus the plane of polarisation is rotatable in discrete steps of 45^{o}.

The heart of the polarimeter is a 40 mm square Wollaston prism which resolves the light into two orthogonally polarised components. The prism consists of two wedges of quartz, with their optic axes orthogonal, cemented together as in figure 2.5, where the action of the prism is illustrated. The emergent light is split into two beams with equal deviations and polarisations which are parallel and perpendicular to the plane of deviation. The prism was manufactured by the Bernard Haller Optical Company and has a nominal deviation of 1^{o}. It is mounted immediately behind the half wave plate in a holder which permits it to rotate for alignment purposes, but with locking screws to hold it fixed when in use.

Behind the Wollaston prism is an electrically

operated shutter which makes a light-tight seal over the detector end of the polarimeter, protecting the detector against accidental excess light. The shutter is provided with an automatic timer for exposures of 4 seconds or less, longer exposures being timed manually.

The final component of the system is a relay lens which, in conjunction with the field lens, re-images the telescope image on to the detector. This is a bloomed f-1.2 Nikon camera lens of 50 mm focal length mounted in a slide with a rack and pinion for focussing accurately on the detector. The overall effect of the field and relay lens combination is to de-magnify the telescope image by a factor of 3.6, resulting in a final image $\sim$10 mm across and a reduction in exposure time by a factor of $\sim$13, necessary when observing faint objects.

The surfaces of the optical components are coated to reduce reflections at all air-glass interfaces.

The operation of the polarimeter is shown in figure 2.3, where it is seen that the 50% of the image admitted by the grid assembly is imaged twice on to the detector in orthogonal polarisation states. The dimensions are so chosen that these two images interlace with the minimum of overlap. Thus the two polarisation components are recorded simultaneously and in close proximity on the detector - an important factor in reducing systemmatic errors.

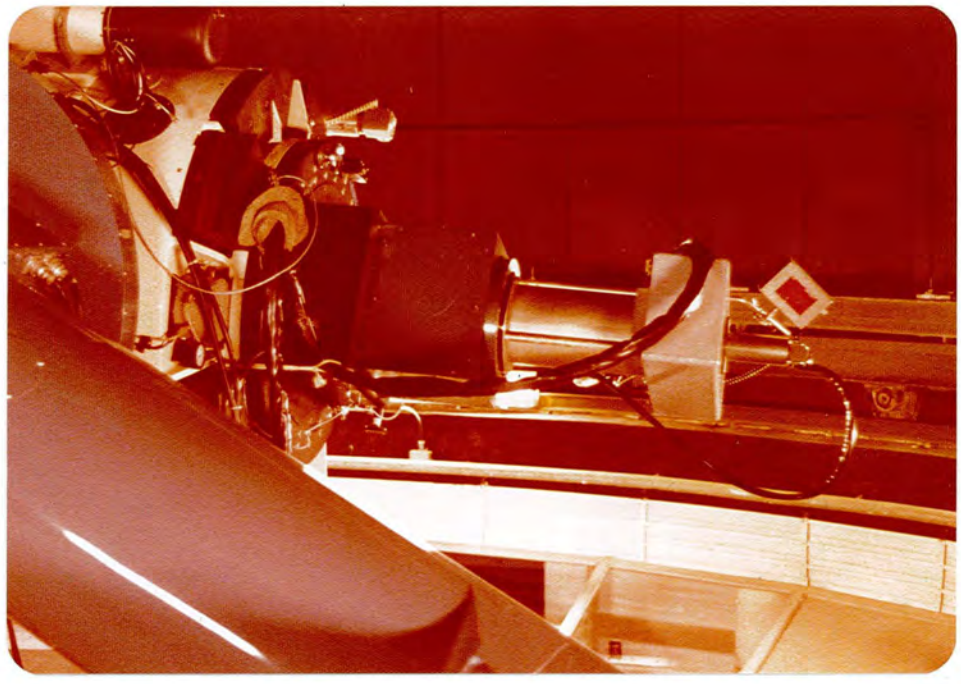
2.3.2 Mechanical Construction

The optics of the polarimeter are housed in a frame constructed of welded, square cross-section steel tubes with 12.5 mm thick aluminium plates bolted to front and back, accurately machined to be perpendicular to the optical axis, and equipped with bolt-holes for attachment to the telescope and electronographic camera.

The relay lens, shutter, prism, half wave plate and filters are mounted on the back plate, and the grid assembly and field lens are rigidly supported by rods running parallel to the axis between front and back plates of the polarimeter. All the optics are easily removable when transporting the instrument.

The rigidity of construction is necessitated by the load of 66 kg produced by the electronographic camera, which must be supported without any flexure which could cause distortion of the recorded images. The steel framework is made light-tight by a covering of aluminium panels, two of which support the control knobs and switches, and two of which are easily removable for internal adjustment. Figure 2.6 shows the polarimeter and electronographic camera installed on the 1 m telescope of the Wise Observatory, Israel.

For use on telescopes which do not have separate facilities for guiding, a detachable guiding box can be fitted between the polarimeter and telescope. This is illustrated in figure 2.7 and consists of a 45° silvered



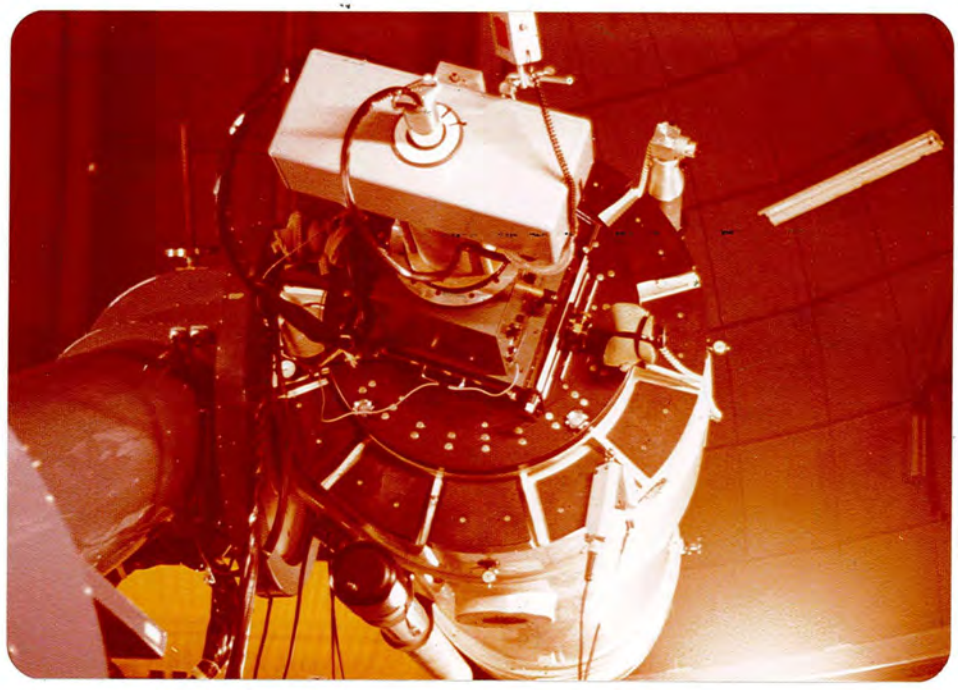


Fig. 2.6 The polarimeter and electronographic camera in use

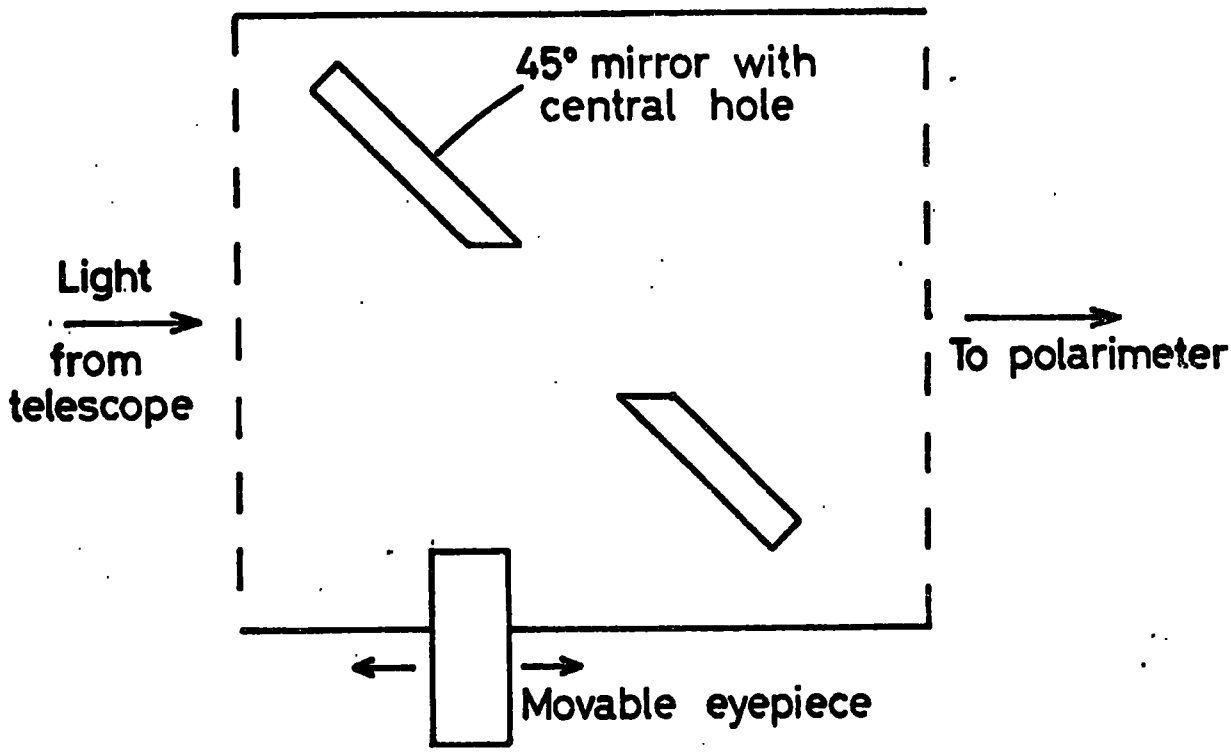


Fig.2.7 The detachable guider box

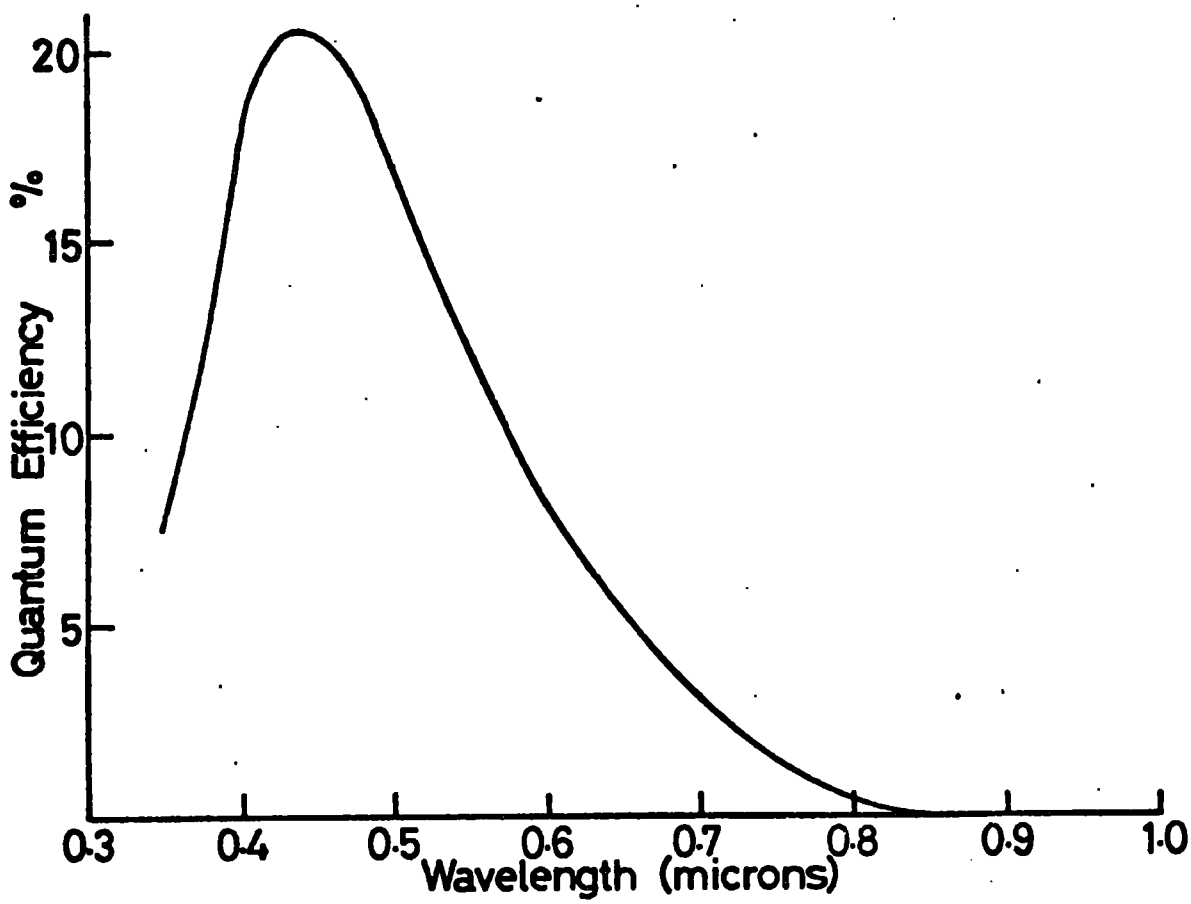


Fig. 2.8 Spectral response of the McMullan camera

perspex mirror with an elliptical hole in its centre, and an eyepiece which can be scanned in 2 dimensions to find a suitable off-axis star on which to guide.

2.4 The Electronographic Camera

2.4.1 Electronography

The electronographic camera, which, apart from early photographic trials, has formed the detector for the Durham polarimeter, was developed at the Royal Greenwich Observatory by Dr. D. McMullan and his colleagues (McMullan et al. 1972). The principle of electronography and its advantages have been known for a long time but it is only recently that a practical device which is easy to use as a general purpose detector for astronomy, has been produced.

The basis of electronography is the conversion of an optical image into an electron image by the photoelectric effect and the subsequent conversion of the electron image into a permanent record by allowing the electrons to strike a photographic plate and produce developable tracks in the emulsion. The advantages resulting from the introduction of an electron stage into the usual photographic process make this form of detector very attractive for astronomy:-

i) The sensitivity is high; the detector quantum efficiency of the electronographic camera (the fraction of the incident photons which produce developable tracks in the emulsion) can be as high as 20% at a wavelength

of 450 nm., compared to about 1.5% for II-ao photographic plates.

ii) The spectral response is determined by the photocathode material, and the detector can be used over a wide wavelength range. A typical response is shown in figure 2.8.

iii) The emulsions used can be fine-grained nuclear emulsions, resulting in a high storage capacity i.e. a wide dynamic range. These emulsions also have low sensitivity to light, enabling them to be handled easily using yellow safelights.

iv) The electrons are accelerated so that they all have sufficient energy to create developable tracks in the emulsion. Thus there is no reciprocity failure as with photographic techniques and coupled with the lack of saturation in fine-grained emulsions, the process is approximately linear over a wide range of typically 0 - 5 in photographic density.

v) The resolution of the camera is superior to that of straight photography, typically 50 - 100 line-pairs/mm., compared to about 30 lp/mm. for II - ao plates.

The main practical problem with the technique arises because the photocathode and electron paths must be maintained in a high vacuum, which is inconsistent with the need to replace the emulsion after each exposure. Early attempts to overcome this problem used complex techniques for introducing emulsions and replacement photocathodes into high vacuum enclosures. The practical

difficulties of using such devices, however, made them of limited use.

The problem was overcome by the introduction of a thin mica window between the high vacuum enclosure and the emulsion. If sufficiently thin this allows electrons to pass through into the emulsion while maintaining the vacuum inside. The difficulty with this solution is that in order to withstand the force of atmospheric pressure on the outside, the mica window must be kept small. This approach has been adopted in the Spectracon tube (McGee et al., 1969) which uses a window 25 x 15 mm. in size.

An alternative approach is to maintain a permanent partial vacuum on the outside of the mica window, thus allowing its size to be increased. The emulsion must then be introduced into this partial vacuum, but this can be done more easily since it does not involve the problems of photocathode contamination and emulsion outgassing present in earlier designs.

2.4.2 The McMullan Camera

This approach has been adopted with considerable success in the McMullan electronographic camera, a schematic diagram of which is shown in figure 2.9. Light enters through the fused silica faceplate and strikes the photocathode which consists of a thin uniform layer of Na, K, Cs and Sb, and is 44 mm. in diameter.

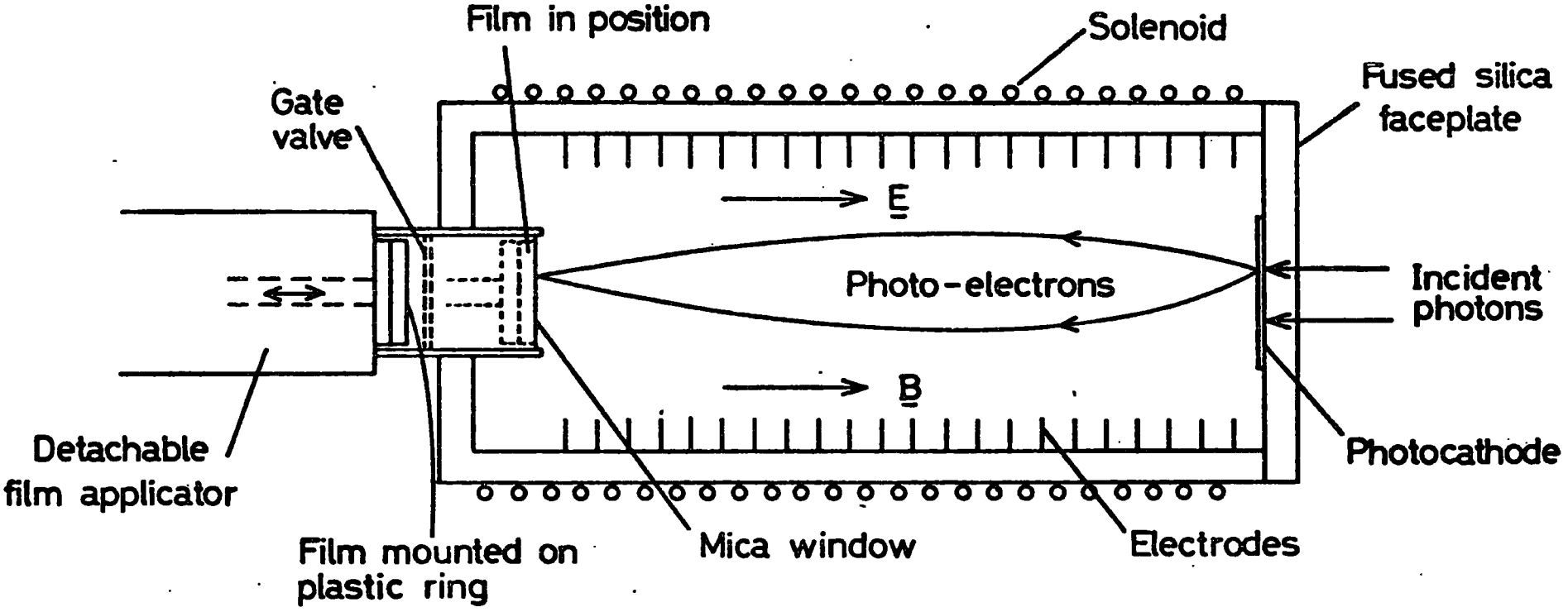


Fig. 29 Schematic diagram of the McMullan camera

Photo-electrons are emitted from this surface with energies of a few eV and are then accelerated down the tube by a uniform electric field along the tube axis produced by a series of electrodes within the tube and a total potential difference of 40 kV. A parallel magnetic field is supplied by a solenoid wound around the tube, and the electron paths become spirals which intersect at a focal plane at the far end of the tube. Focussing is accomplished by varying the solenoid current, and the tube vacuum is maintained by an ion appendage pump.

The mica window, 40 mm. in diameter and 4μm. thick, is set at the focal plane. The electrons are scattered in passing through the window, and about 30% are lost, but the emulsion is maintained in close contact with the outer surface of the mica and little resolution is lost as a result. The electrons then stop in the emulsion producing tracks which are later developed to produce a negative picture of the original optical image. The process of loading and changing films through a vacuum lock is as follows:-

The film, mounted on a plastic ring, is clipped on to the applicator, which is inserted into the back of the camera making a vacuum seal. The space above the film is then evacuated, and when this is complete a gate valve is opened to allow access to the mica window. The applicator then drives the film forward against the window, and air is applied at a pressure of $\sim$10 Torr to

the back surface of the film, pressing it forward to make close contact with the mica. An exposure is then made. To remove a film the applicator is driven back, the gate valve shut and air admitted above the film. The applicator is then removed and a new film clipped in place.

The whole electro-pneumatic process is controlled by a free-standing programmer unit which includes interlocks to prevent accidental damage to the tube. In practice a film can be changed in under 2 minutes.

2.5 Polarimeter Alignment

For protection, all the optics are removed when transporting the polarimeter, so that alignment must be performed at the observing site. This is done before mounting the instrument on the telescope, as follows:-

i) With the filter and half wave plate omitted, an eyepiece is positioned behind the relay lens so that the grid assembly can be viewed. Two images are seen as described previously. A straight edge is positioned across the grid assembly, perpendicular to the grid bars, and the prism is rotated until the straight edge appears continuous across the two images (Figure 2.10). This ensures that the prism deviates the light in a plane perpendicular to the grid bars. These then serve as a reference direction for the whole instrument.

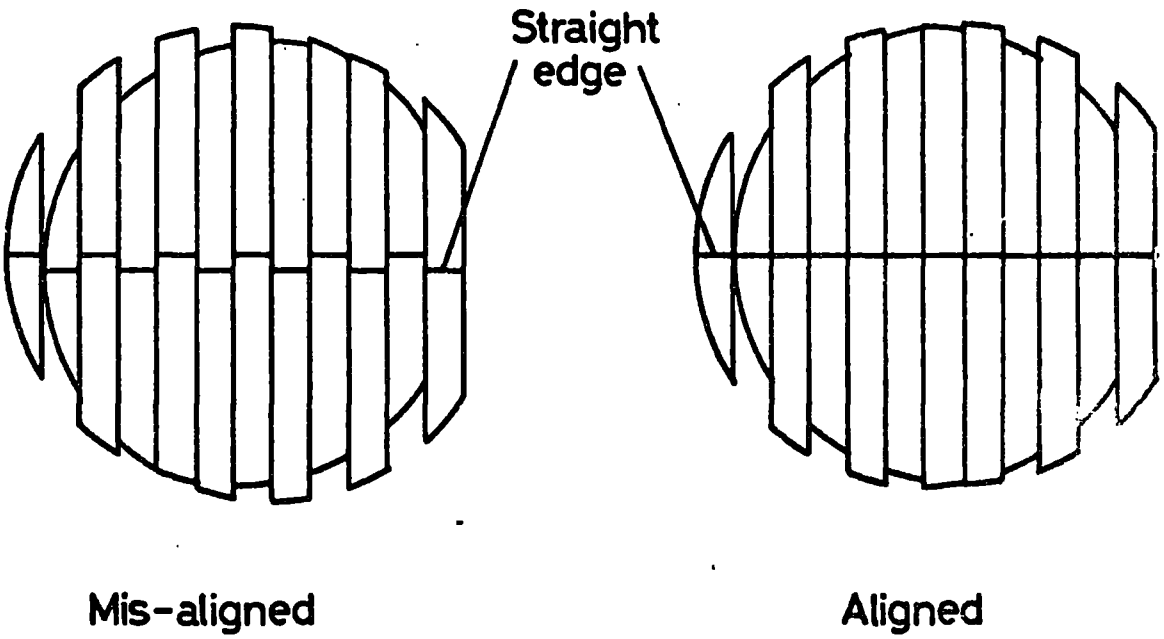


Fig. 2.10 Aligning the Wollaston prism

ii) A polaroid sheet is then inserted in front of the grid assembly and, using a bright light source, is rotated until one of the images of the grid is extinguished - the preferred direction of the polaroid is then parallel to the grid bars.

iii) The half wave plate and filter are inserted, and the half wave plate adjusted so that at one of its preset positions the same image of the grid assembly is extinguished. The optic axis of the half wave plate is then set parallel to the grid bars and alignment is complete. The polaroid is then removed.

Since the grid bars serve as a reference direction, the orientation of the polarisation on the sky can be determined by comparing this direction with images of known astronomical features (e.g. stars) on the final image. Thus it is not necessary to align the polarimeter in any fixed direction on the telescope. For convenience, however, the grid bars are sometimes aligned North-South or East-West. The method of doing this is described in the next section.

2.6 Using the Polarimeter

2.6.1 Starting Up

When the polarimeter and camera are mounted on the telescope, some checks are necessary before observations can commence. The starting up procedure is thus as follows:-

i) The camera control system is powered up, its operation is checked, and it is left running to stabilise.

ii) The telescope is uncovered and a bright star is centred in the polarimeter field.

iii) With the grid viewing eyepiece removed, an image of the primary mirror illuminated by starlight is seen, and this is used to focus the telescope on to the polarimeter grids by using the knife-edge test, the grid bars being used as the knife-edge. Telescope focussing is accomplished by moving the secondary mirror until the image seen darkens evenly all over as the knife-edge is brought into the beam. The viewing eyepiece is then replaced.

iv) The camera is loaded and a polarimeter focus test made by recording several star images on one electronograph, with different settings of the relay lens position. The best image is chosen and the lens set to the corresponding position. The camera magnetic focussing may also be checked by this means if necessary. Perfect focussing can also be judged by observing the images of the grid spots, which should be perfectly circular with sharp edges.

v) If required, the grid bars are set North-South by driving the telescope in declination and rotating the polarimeter on the telescope until stars seen in the plane of the grid assembly appear to move parallel to the grid bars. Other orientations are found by rotating from this

reference position.

2.6.2 Making the Observations

When the above checks have been made and everything is operating, observing can commence. The sequence of operations is as follows:-

i) The object to be observed is positioned in the polarimeter field, ensuring that the field also contains a region of night sky for background subtraction and at least one star for image registration in the data-reduction stages.

ii) A suitable off-axis guide star is chosen and centred in the guiding system.

iii) The grid system is pushed fully in one direction (referred to as IN or OUT) and the grid viewing mirror is removed from the beam.

iv) The camera is loaded with a new film.

v) The half wave plate is set at one of 4 adjacent preset positions (referred to, for historical reasons, as positions 13, 14, 15 or 0) and an exposure made.

vi) Steps (iv) and (v) are repeated for the other 3 half wave plate positions.

vii) Another set of 4 exposures is then made with the grid assembly set to the opposite position to cover the other half of the object.

2.6.3 Calibrating the Detector

In order to calibrate the photocathode of the camera, exposures are made during each observing run using uniform illumination of the field. This is done using the same filters as are used for the observations and 4 exposures are made for each position of the grid assembly.

Uniform illumination is achieved by viewing the inside of the telescope dome, lit from within. This does not result in a spatially uniform brightness on the inside of the dome, however, the telescope is focussed at infinity and so in fact performs angular resolution of the light. The angular uniformity of such a source of illumination appears to be quite adequate for the purpose of calibrating the electronographic camera.

2.7 Film Processing

2.7.1 The Films used in Electronography

The films used with the McMullan electronographic camera consist of a backing layer of Melinex 50μm. thick, supporting a 10μm. thick layer of emulsion. The emulsions used are Ilford L4 and G5 nuclear research emulsions. L4 is a fine grained emulsion resulting in a large dynamic range but visually unrewarding pictures, since all the information is stored in a small range of photographic density which the eye cannot appreciate. G5 is a coarser grained emulsion and gives pictures which are more readily appreciated by eye, giving the impression

of greater sensitivity. However, the noise level is higher and the dynamic range less than L4. The relative merits of the two emulsions will be discussed in more detail in chapter 9.

Since the quality of the film and its processing are relevant to the results obtained during the data-reduction process to be described later, the film handling will be described here.

2.7.2 Film Handling

The film is supplied mounted on glass plates from which it is peeled and cut into discs 46.5 mm. in diameter for use. At this stage inspection is carried out to reject any film which has dirt adhering to it, is scratched, or shows signs that the emulsion is not of uniform thickness. The discs of film are mounted on plastic rings for use, using double sided sticky tape rings. Before loading into the camera, each film is brushed lightly with a camel's hair brush to remove dust, and it is written on for identification on removal from the camera to prevent ink being transferred from the film to the mica window.

The films are developed shortly after exposure using the following sequence:-

i) Pre-soak in distilled water for 2 minutes to soften the emulsion and speed developer penetration.

ii) Develop in full strength D19 developer for 4 minutes using nitrogen bubble-burst agitation.

iii) Stop bath of distilled water and 0.5% acetic acid for 10-15 seconds.

iv) Fix for 4 minutes in Kodak rapid fixer.

v) Wash for 1 minute in distilled water.

vi) Hypo-clearing agent (1 part to 4 parts of distilled water) for 1½-2 minutes.

vii) Wash in running, filtered water for 15-20 minutes.

viii) Rinse in 2 parts distilled water + 1 part ethanol + few drops of Photoflo wetting agent.

ix) Dry in a stream of air, removing drops of liquid with absorbent paper.

x) Store between sheets of soft paper.

All the solutions used during film processing are maintained at 20^{o} C $\pm$ 1^{o} ($\pm$ 0.5^{o} for the developer) and, whenever possible, a whole sequence of 4 electronographs comprising a set of observations are processed together, to minimise any effects due to developing variations.

CHAPTER 3

DIGITISATION

3.1 Photographic Density

The purpose of digitisation is to convert the raw data from the polarimeter, in the form of electronographs, into numerical data which can be handled by computer to produce the final results i.e. a polarisation map.

The quantity which is required from the electronographs is the photographic density or 'blackness' of the image at each point since this is proportional to the number of photons which fell on the photocathode of the camera at the corresponding point during the exposure. Density is defined on a logarithmic scale as follows:-

Suppose a beam of light is shone through the electronograph at a point and a fraction T of the incident light is transmitted, then the density D at that point is defined as:

$$D = -\log_{10} T \qquad 3.1$$

i.e. a density of 1 transmits 10% of the light, D = 2 transmits 1%, etc.

3.2 The PDS Microdensitometer

The machine used to perform digitisation of the polarimeter electronographs is the PDS Scanning Microdensitometer at the Royal Greenwich Observatory. In this machine, light from an incandescent lamp is focussed

through an aperture on to the electronograph which is supported on a glass table. The transmitted light passes through a similar optical system with a slightly smaller aperture, which defines the area of electronograph measured, and falls on a photomultiplier tube. The output from this tube is taken to a logarithmic amplifier to produce a result proportional to the density of the electronograph within the aperture used. The output of this amplifier is taken to an analogue to digital converter to produce digital results which are recorded on magnetic tape.

No reference light beam is used in this machine, which means that the stability of measurement depends on the stability of the electronics, particularly the logarithmic amplifier, and of the light source.

3.3 The Scanning Process

The region of electronograph measured is determined by moving the glass table supporting it. This movement is computer controlled in two directions, positions being determined from optical encoders giving a positional accuracy of about $\pm$ 0.2μm. For the polarimeter electronographs, measurements are made by sampling the density on a 512x512 matrix of points spaced by 25μm. in the X and Y directions and covering the entire polarimeter image and some of the unexposed film. This is accomplished by means of a raster scan, illustrated in figure 3.1.

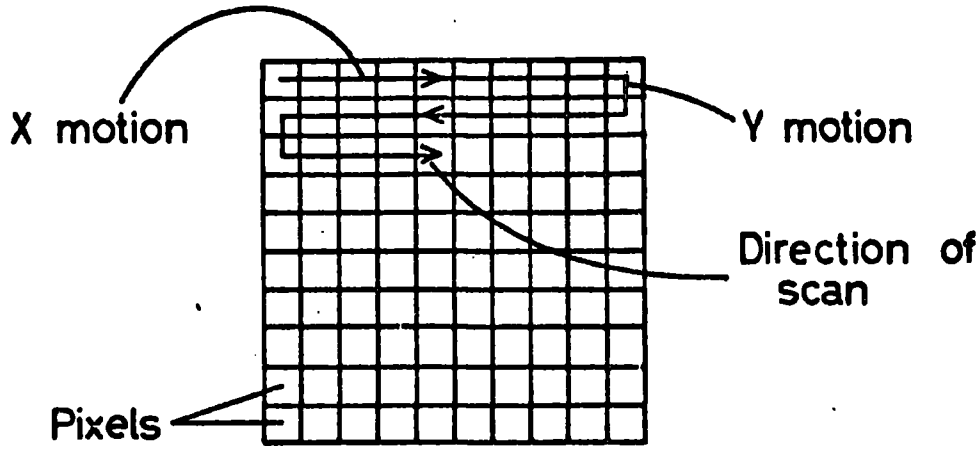


Fig. 3·1 The raster scan used

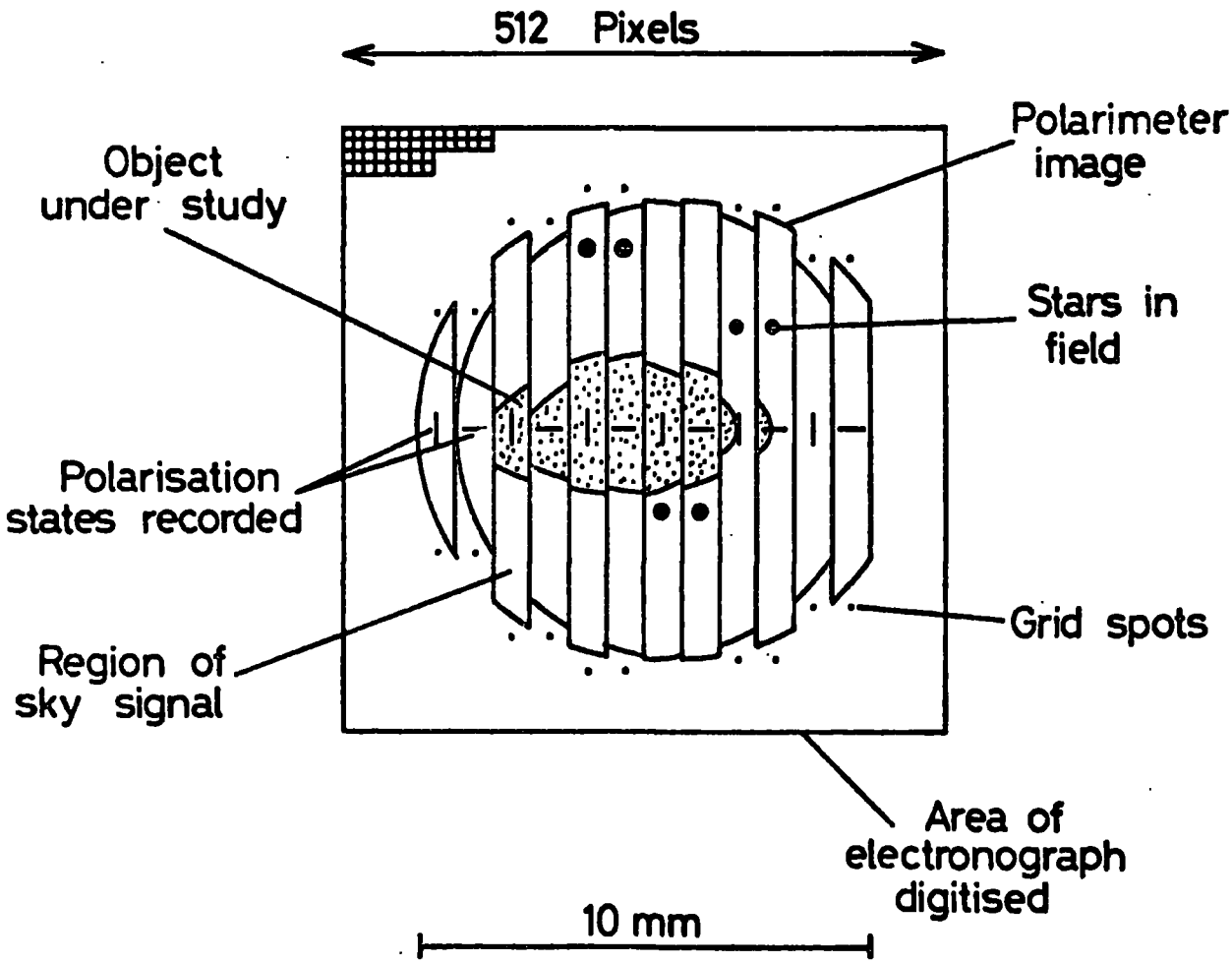


Fig. 3·2 The regions of an electronograph digitised

The table moves along scan lines in the X direction, the machine sampling the density within the aperture, which in the case of the polarimeter electronographs is 24μm square. Every 25μm of travel, a measurement is made and at the end of a scan line, the entire line of 512 measurements is written on to magnetic tape. The scan then continues on the next line, incremented by 25μm in Y, in the opposite direction, the process continuing for 512 lines. In this way the image is split into 262, 144 picture elements or "pixels".

The raster scan is used as it is most efficient as regards time. The scan speed is determined by the speed of response of the PDS measuring system, which must respond to changes in density as the image is scanned through the beam. The response time increases when high densities are being measured and this means that the scan time is dominated by the need to measure small, dense regions of the film, such as star images. In order that no significant distortion of star images should occur, a total scan time of $\sim$35 minutes per electronograph is required. Some time saving is possible in the case of those electronographs obtained using uniform illumination of the field for calibration purposes. Since these do not contain small, dense regions they may be scanned in $\sim$10 minutes without ill effect.

Before scanning commences, the electronographs are

aligned on the table of the PDS machine with the direction of the grid bars parallel to the Y axis, and the grid spots are used as reference marks to ensure that approximately the same area of each electronograph is scanned. the appearance of an electronograph, the region scanned and the features relevant to the data-reduction procedure are illustrated in figure 3.2.

3.4 The Numerical Data

The numbers representing density produced by the PDS machine are integers in the range 0 - 1023 and it is these numbers which are written on to magnetic tape for computer processing. The conversion from these integers to density is approximately such that a change in output of 200 corresponds to a change in density of 1, so that the total density range spanned is from 0 to 5 with a quantisation interval of 0.005 in density. The integers output by the machine are referred to for convenience as "PDS units" throughout this thesis.

It will emerge that a typical accuracy with which measurements can be made is better than 1 PDS unit. It might thus be thought that a smaller quantisation interval would be necessary. However, on detailed investigation, it appears that both the quantisation interval, and the spacing of sampled points on the electronograph are near to the optimum values for this type of work. Precisely

why this is so requires a fuller discussion of the type of errors present, and this will be left until chapter 7.

The magnetic tapes produced by the PDS machine are 7 track tapes, each capable of holding the data from 11 electronographs. The first step in processing the data is to move it to a 9 track tape in a format which can be read at the computing centre where data reduction takes place. This conversion is performed at the Royal Greenwich Observatory and the tapes are then sent to the University of Newcastle computing centre, where they can be accessed at computer terminals at Durham University.

The data is transferred to magnetic disc storage for use, alternate lines of each scan being reversed to undo the effect of the raster scan on the PDS machine, and figure 3.3 shows a computer line-printer picture of the data taken straight from disc storage. The data is condensed from a 512x512 matrix to a 128x128 matrix, each value being the average of 16 measurements, and the value at each point is represented by one of the characters: blank, 123456789ABCDEFGHIJKLMNOPQRSTUVWXYZ @%+*, in order of increasing density, each character representing a range of 25 PDS units. The clear, unexposed film, the outline of the polarimeter field and the object and stars within the field are clearly seen.

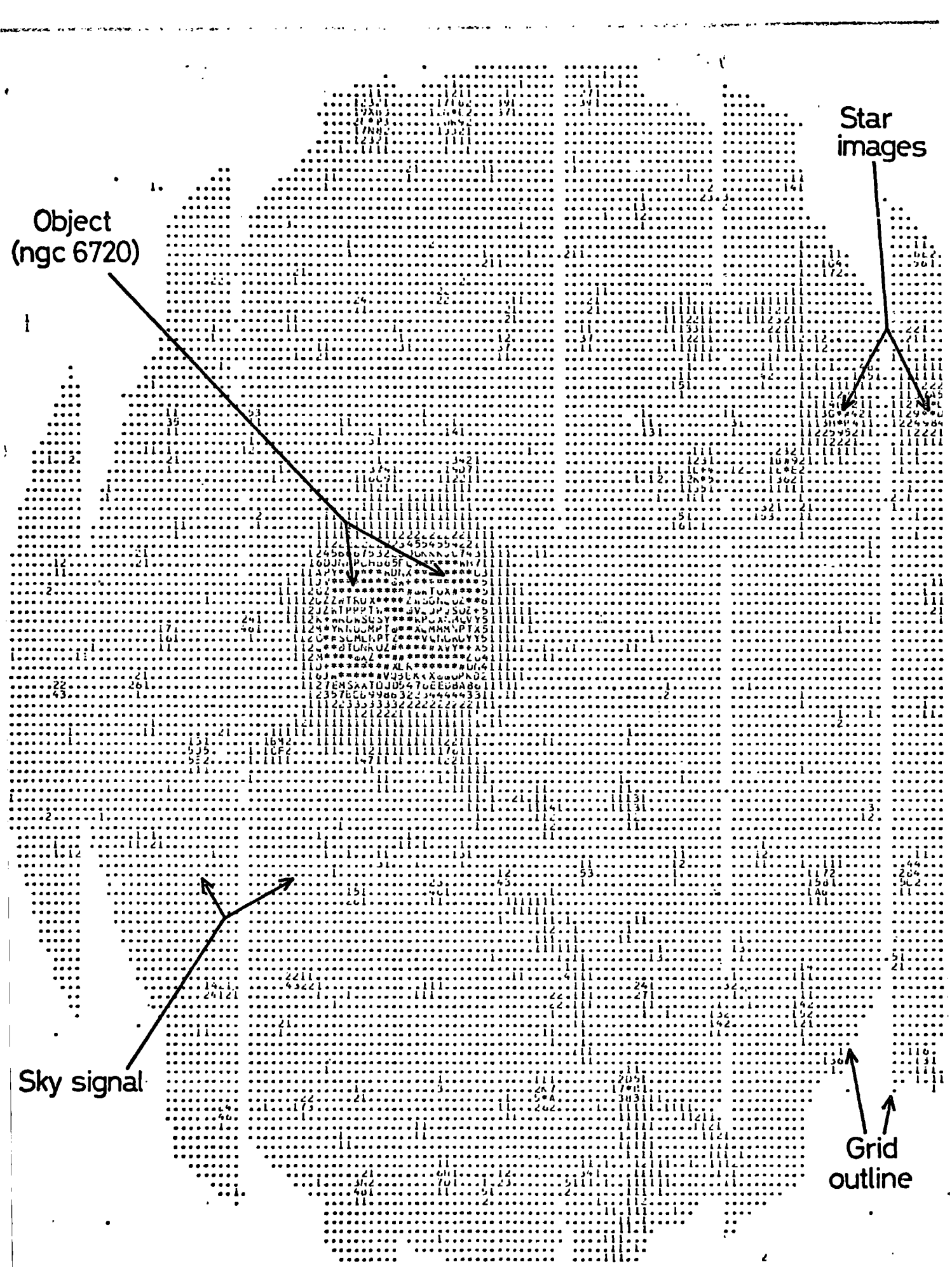


Fig.3.3 A line printer picture of the raw data

CHAPTER 4

INTRODUCTION TO THE REDUCTION TECHNIQUE

4.1 Simple Theory of Data Reduction

The data obtained from the polarimeter consists, ideally, of 8 images of the object under study. These are obtained in 4 pairs, the images in each pair being resolved into orthogonal linearly polarised components with the direction of polarisation in each pair being rotated by 45^{o} with respect to the previous pair. This is illustrated in figure 4.1. We represent the densities at corresponding points on the 8 images by d_1 - d_8. These densities are related in a simple way to the light intensities which produced them, denoted by i_1 - i_8, namely:

$$d_1 = i_1 \qquad 4.1$$

$$d_2 = Fi_2 \qquad 4.2$$

$$d_3 = E_1 i_3 \qquad 4.3$$

$$d_4 = E_1 F i_4 \qquad 4.4$$

$$d_5 = E_2 i_5 \qquad 4.5$$

$$d_6 = E_2 F i_6 \qquad 4.6$$

$$d_7 = E_3 i_7 \qquad 4.7$$

$$d_8 = E_3 F i_8 \qquad 4.8$$

Here, the constant of proportionality for d_1 has been set to unity, F represents the relative sensitivity of the system when recording the right hand image with respect to the left hand image (remembering that they

Electronograph number	Polarisation states	Orientations	Densities recorded
1		0° 90°	d_1 d_2
2		45° 135°	d_3 d_4
3		90° 180°	d_5 d_6
4		135° 225°	d_7 d_8

Fig. 4-1 The polarisation states recorded

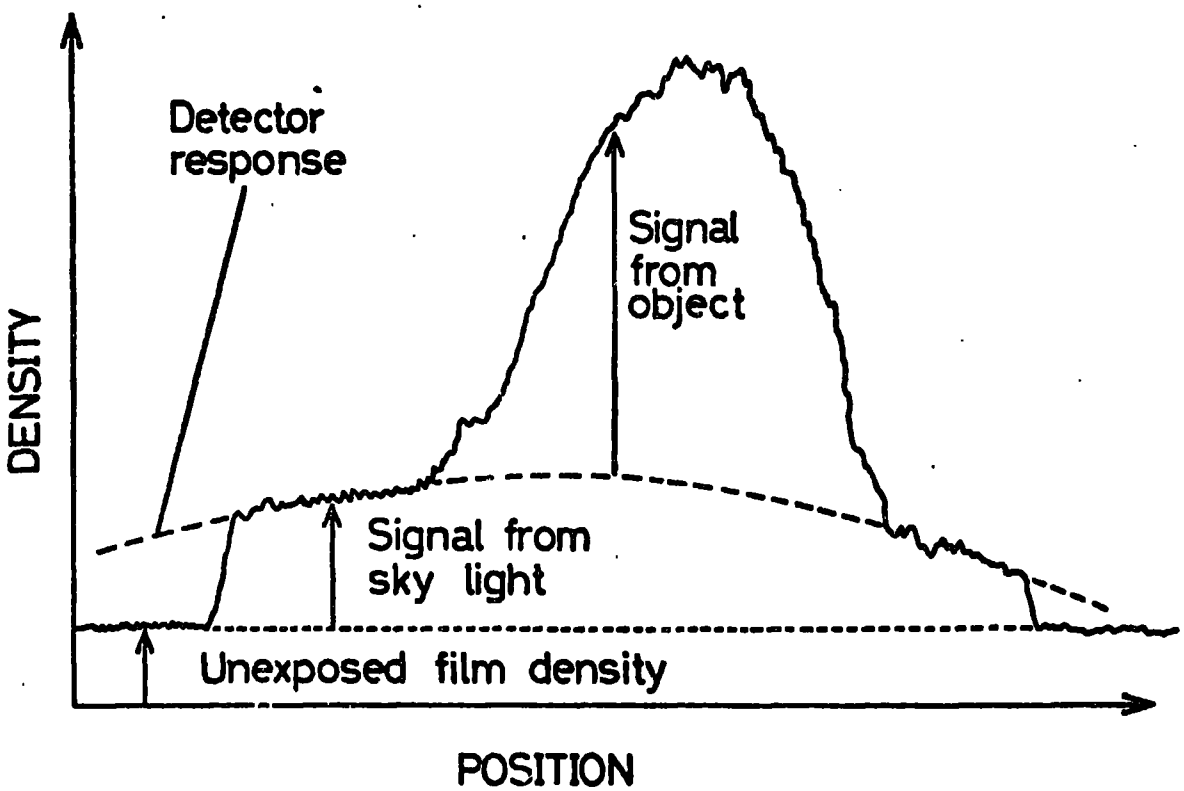


Fig. 4-2 The signals recorded on an electronograph

are recorded by different parts of the photocathode and have different optical paths in the polarimeter), and E_1, E_2, E_3 are "exposure factors" representing light lost for any reason e.g. changing atmospheric extinction, slightly different exposure times, etc. In practice, F, E_1, E_2, E_3 all lie close to unity.

In addition, we have the following relations between the light intensities because of the orientations of the polarisation states which we chose to use (see figure 4.1):

$$i_1 = i_6 \qquad 4.9$$

$$i_2 = i_5 \qquad 4.10$$

$$i_3 = i_8 \qquad 4.11$$

$$i_4 = i_7 \qquad 4.12$$

$$i_1 + i_2 = i_3 + i_4 = i_5 + i_6 = i_7 + i_8 \qquad 4.13$$

From these relations (4.1 - 4.13), we can calculate the scale factors F, E_1, E_2, E_3. In fact we have two estimates of f:

$$F_1 = \sqrt{\frac{d_2\ d_6}{d_1\ d_5}} \qquad 4.14$$

and

$$F_2 = \sqrt{\frac{d_4\ d_8}{d_3\ d_7}} \qquad 4.15$$

Taking a mean value:

$$F = \tfrac{1}{2}(f_1 + f_2) \qquad 4.16$$

we can now find:

$$E_1 = \frac{d_3 + d_4 / F}{d_1 + d_2 / F} \qquad 4.17$$

$$E_2 = \frac{d_5 + d_6 / F}{d_1 + d_2 / F} \qquad 4.18$$

$$E_3 = \frac{d_7 + d_8 / F}{d_1 + d_2 / F} \qquad 4.19$$

Knowing these scale factors we can derive $i_1 - i_8$ from the measured densities $d_1 - d_8$ and calculate the Stokes parameters using the following relations:

$$I = \tfrac{1}{4} (i_1 + i_2 + i_3 + i_4 + i_5 + i_6 + i_7 + i_8) \qquad 4.20$$

$$Q = \tfrac{1}{2} \left[(i_1 - i_2) + (i_6 - i_5)\right] \qquad 4.21$$

$$U = \tfrac{1}{2} \left[(i_3 - i_4) + (i_8 - i_7)\right] \qquad 4.22$$

These relations follow directly from the definitions of the Stokes parameters (1.2), and express the fact that I is the total light intensity, Q the preference of the light for the 0^o orientation and U its preference for the 45^o orientation.

The only simple error made in this analysis is the

neglection of the night sky signal. This must be subtracted from the densities d_1 - d_8 before the analysis is performed. Since, however, the sky density must be sampled from a region of the field different to the region where it is to be subtracted, it is necessary to know the relative sensitivities of the detector in these two regions, i.e. the spatial response of the detector must be corrected for (Figure 4.2).

This, then is the simple theory behind the polarimeter. Before this can be applied it only remains to align the images on the electronographs with each other so that identical regions of the object are used for each density d_1 - d_8. This can be done by aligning the stars on the separate images and this process will be described more fully in the next chapter. Figure 4.3 shows a block diagram of this simple reduction technique.

4.2 Limitations of the Electronographic Technique

4.2.1 Introduction

While the electronographic detection system has many features which make it superior to photography when numerical results are required, it also has a number of limitations which mean that a practical data reduction system must be considerably more complex than the above analysis would suggest. Some of the important effects which complicate the reduction are detailed below.

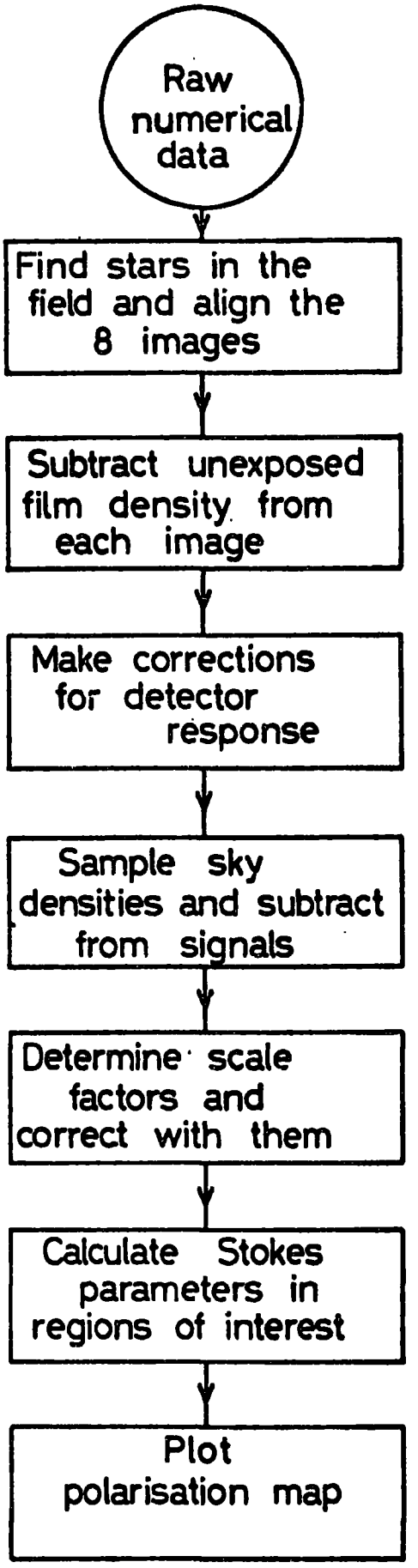


Fig. 4.3 A simple reduction technique

4.2.2 Emulsion Linearity

The photographic density on an electronograph is not strictly proportional to the density of electron tracks, as the following analysis shows:-

Consider a film of thickness Δt and unit area, which contains N grains per unit volume and suppose that an electron passing through this thickness of film renders $sN\Delta t$ grains developable.

Since there are a finite number of grains available, we must allow for the fact that subsequent electrons will not be able to activate as many because a certain proportion, say n per unit volume will already have been made developable by previous electrons. Hence the number of grains activated per electron will be reduced to $s(N - n)\,\Delta t$, and a change in electron density de' is related to the change in developable grain density dn by:

$$S(N - n)\,\Delta t\, de' = \Delta t\, dn \qquad 4.23$$

Integration then leads to the relation:

$$n = N(1 - \exp\,[-se']) \qquad 4.24$$

The fraction T of light transmitted by such a thickness of film after development is:

$$T = 1 - na\Delta t = 1 - Na\Delta t\,(1 - \exp\,[-se']) \qquad 4.25$$

where a is the cross-sectional area of a developed grain.

If the film is of finite thickness t, considered as $t/\Delta t$ separate layers each of thickness Δt, then the fraction transmitted is:

$$T = \lim_{\Delta t \to 0} \left\{ 1 - Na\,\Delta t \,(1 - \exp\left[-se'\right]) \right\}^{t/\Delta t} \qquad 4.26$$

$$= \exp - \left\{ Nat \,(1 - \exp\left[-se'\right]) \right\} \qquad 4.27$$

Hence the measured photographic density is:

$$D_M = -\log_{10} T = Nat \,(1 - \exp\left[-se'\right]).\ \log_{10} e \qquad 4.28$$

This is thus <u>not</u> proportional to the electron density.

If, instead, we consider the number of grains in the film to be very large and neglect the density n of grains already rendered developable when we calculate the effect due to further electrons, we arrive at a 'true' photographic density which <u>is</u> proportional to the electron density:

$$D_T = Nat.\ se'.\ \log_{10} e \qquad 4.29$$

Eliminating the electron density between equations 4.28 and 4.29 we can relate the true density to that measured:

$$D_T = -\,Nat.\ \ln\,(1 - D_M/Nat) \qquad 4.30$$

The constant Nat has been determined, for the case where the density is measured in PDS units, to be 1890 and 3780 for G5 and L4 emulsions respectively, these figures representing the maximum or saturation density which can be obtained with the emulsions. In practice, these values are well above the range normally used so that the corrections which need to be applied are small. However, as will be described in a later section on the quality of electronographic film, there is some reason to believe that the film parameters, on which these constants depend, may vary (9.5.1). Consequently the corrections applied may be slightly in error in some cases.

4.2.3 Clear Plate Subtraction

The unexposed part of the electronographic film has a non-negligible photographic density (referred to as the "clear plate signal") (Figure 4.2). Problems can arise in determining this signal to the required accuracy, especially as it may not be constant over the electronograph (5.1.1).

4.2.4 Effects Due to the PDS Machine

The response of the PDS machine to photographic density is not stricly linear, and linearity corrections must be applied to the measurements made on it. (5.1.6)

A more serious problem, however, is the lack of stability of the microdensitometer. Numerous attempts have been made by the staff of the Royal Greenwich Observatory to improve this, but without complete success. From a user's point of view, this instability is apparent as a slow drift of zero level during a scan, sometimes accompanied by abrupt jumps. The amplitude of these can reach tens of PDS units, so some form of zero level monitoring is essential. (5.1.2) It is probable that these jumps remain a serious source of error.

4.2.5 Bad Data

By far the most serious problem in the reduction technique is the occurence of bad data, i.e. measurements which are corrupted by blemishes on the electronographs. These blemishes can take many forms, due for example to "dead" spots and "hot" spots on the photocathode, contamination of the mica window, dirt and scratches on the film occurring before or after processing, defects in the emulsion and drying stains. While every effort is made to prevent these blemishes, some contamination occurs on every electronograph. Figure 4.4 shows examples of this.

Because of this problem of corrupt data, tests must be performed at nearly every stage of the data reduction technique to reject it before any quantities are calculated which could be affected by its presence. Chapter 7 describes

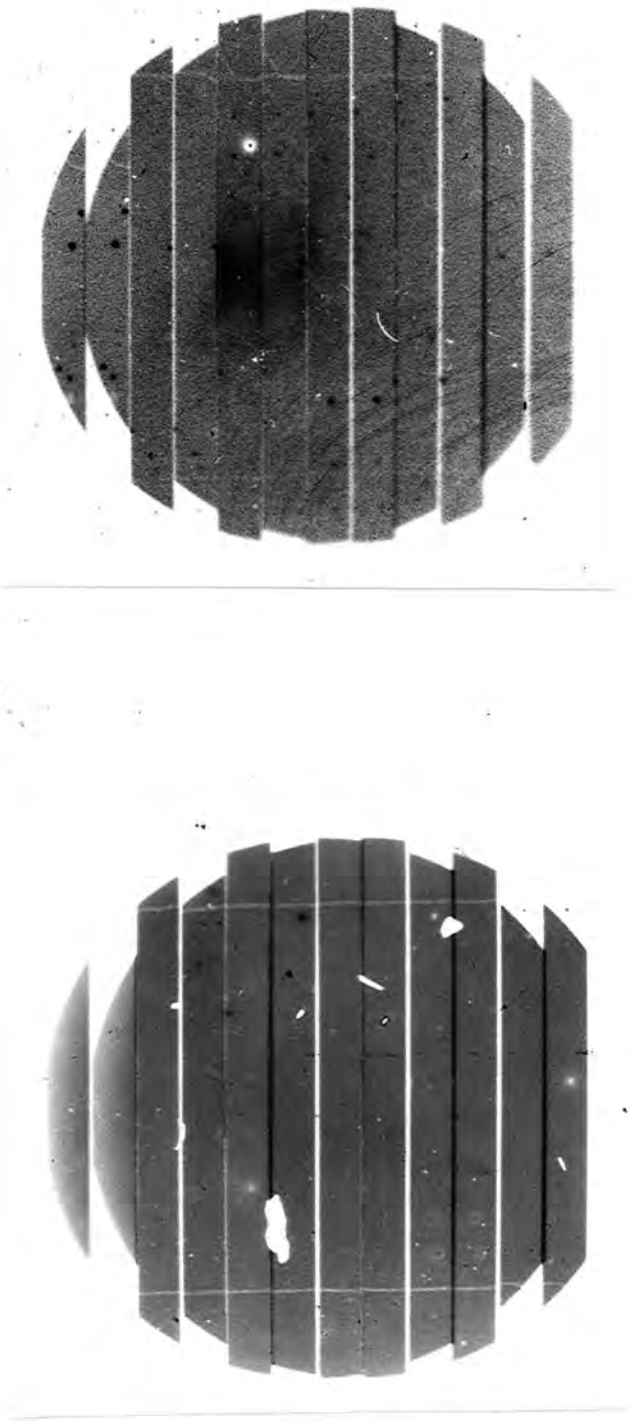


Fig.4-4 Examples of dirt, scratches and a 'dead spot' (above) on the film and (below) particles on the mica window

an effort to deal with this.problem as effectively as possible before the final polarisation parameters are calculated.

4.3 Development of the Data Reduction Technique

The new data reduction technique to be described was developed by the author in conjunction with Dr. W. S. Pallister of the University of Durham. The development grew out of an attempt to locate the sources of error in an existing reduction technique (Axon, 1977; Pallister 1976), based on the simple outline given at the start of this chapter.

In attempting to allow for some of the limitations of electronography it became apparent that a basically more careful approach was needed to the problem of extracting numerical results from electronographs if the full potential of the technique was to be exploited. The new technique includes not only an attempt to identify and correct for a number of previously unrecognised sources of error, but also improvements to a number of standard techniques, such as photocathode calibration and clear plate subtraction, which provide more accurate results than previous methods, particularly in the presence of corrupt data. It also includes a major re-structuring of the data-reduction programmes in an attempt to introduce automation, eliminating much tedious interactive computer work and the associated possibility of human error, and

allowing decisions at various stages to be made by the computer on an objective basis, resulting in consistently reproducible results.

One of the most useful aspects of the new technique is that it allows continual monitoring of errors at each stage of the reduction process, and an examination of the internal consistency of the data. This has been used to provide a new and efficient method of rejecting corrupt data and of determining the probable errors on the resulting polarisation parameters from a variety of causes. With this knowledge it has been possible to identify the poor quality of electronographic film as the major source of error, and it is hoped that if suitable improvements can be made, electronography may compare favourably with other modern detecting devices.

4.4 Computing Considerations

The computer programmes required in the data reduction procedure have been written in the Fortran IV programming language and all the computer work has been done using interactive terminals at the University of Durham computer unit connected to the Northumbrian Universities Multiple Access Computer (NUMAC) in the University of Newcastle computing centre.

The machines used in the NUMAC system are an IBM 360/67 and 370/168. These use a "virtual memory"

system by which the contents of the computer core are written on to a magnetic drum when not required without the user being aware of this. The result is to simulate a very large storage space in the computer core which can be accessed via programmes. This facility means that the large amounts of data involved in the computations to be described can be loaded directly into the computer core and space restrictions are virtually absent. To make economical use of the storage space, however, the data is stored as half-word integers whenever possible (INTEGER*2 in Fortran), thus halving the amount of storage space required. Extensive use has also been made of a 200 Megabyte magnetic disc and magnetic tapes for long-term data storage.

CHAPTER 5

EXTRACTING NUMERICAL DATA FROM THE ELECTRONOGRAPHS

5.1 Clear Plate Subtraction

5.1.1 Introduction

The first step in the data reduction procedure is to subtract the clear plate density from the data, as this represents the zero level upon which all the signals are superimposed (Figure 4.2).

Figure 5.1 shows a typical clear plate density (The symbols are as in figure 3.3, each representing a range of 0.5 PDS units, and are chosen so as to show the unexposed region of the electronograph only). As can be seen, the density is not constant and some form of interpolation is required in order to subtract a realistic quantity from the area of signal. The method adopted is linear interpolation between the ends of each scan line. First, however, we must sample the density at the end of each line and obtain a realistic result, as any errors in this small region will affect an entire line of data.

5.1.2 Sampling the Clear Plate Density

The first step is to select a region of unexposed film to sample. The region used is shown in figure 5.1,

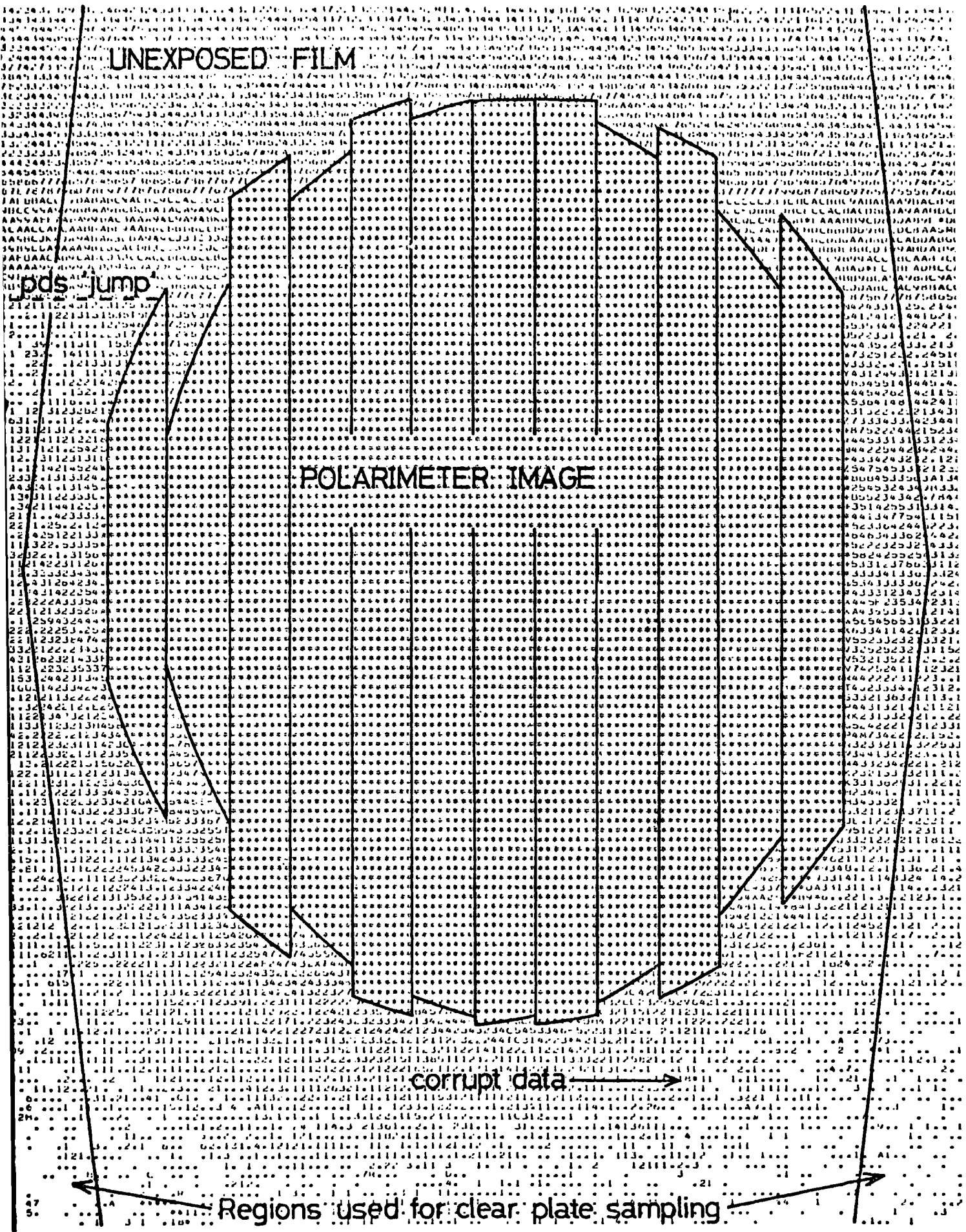


Fig. 5.1 Typical clear plate signal

the number of pixels used in each line being typically 20 to 50. A histogram of densities in this region is formed at each end of each line, and the density bin between 0 and 60 PDS units which contains the maximum number of pixels is found (The clear plate density is known to lie in this range). An average density d is then found from the 5 density bins centred on this value and an error σ estimated from the width of the histogram and the number of pixels used. In this way, pixels with densities far from the mean value are not used (Figure 5.2).

Repeating this process for all 512 lines in a scan yields 512 pairs of samples of the density and associated estimates of the errors. Similar methods of deriving these have been used by other workers. The problem with the method is demonstrated in figure 5.3, which shows the values obtained on a typical electronograph, at one end of each line, plotted against the line number within the scan. As can be seen, the graph is contaminated by lines where the above procedure has failed to reject all erroneous pixels, probably because the majority of pixels in these lines were erroneous. This effect occasionally extends over several lines. In addition, drift and sudden jumps due to the PDS machine are often encountered and can reach serious proportions; figure 5.4 shows an example of this. Such effects show simultaneously at both ends of each line.

Because of the relatively small number of pixels

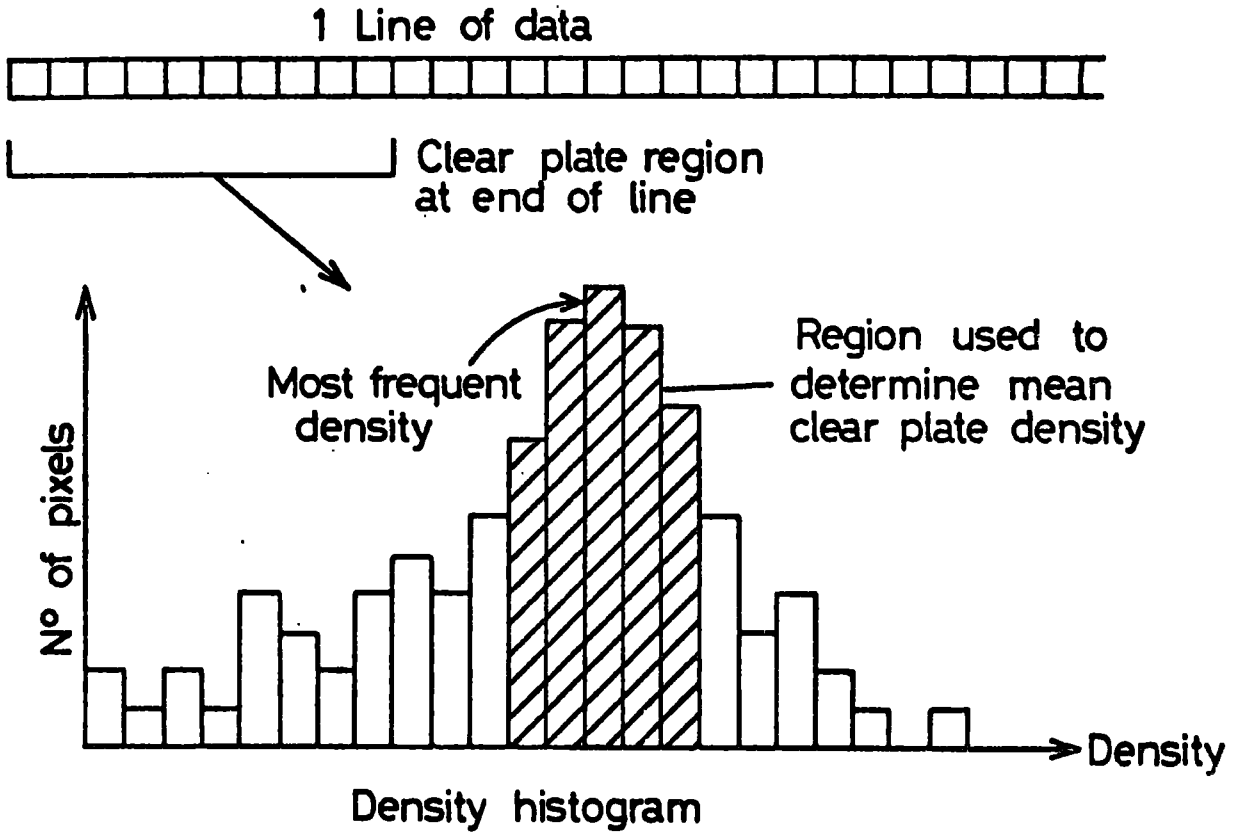


Fig. 5.2 Sampling the clear plate density

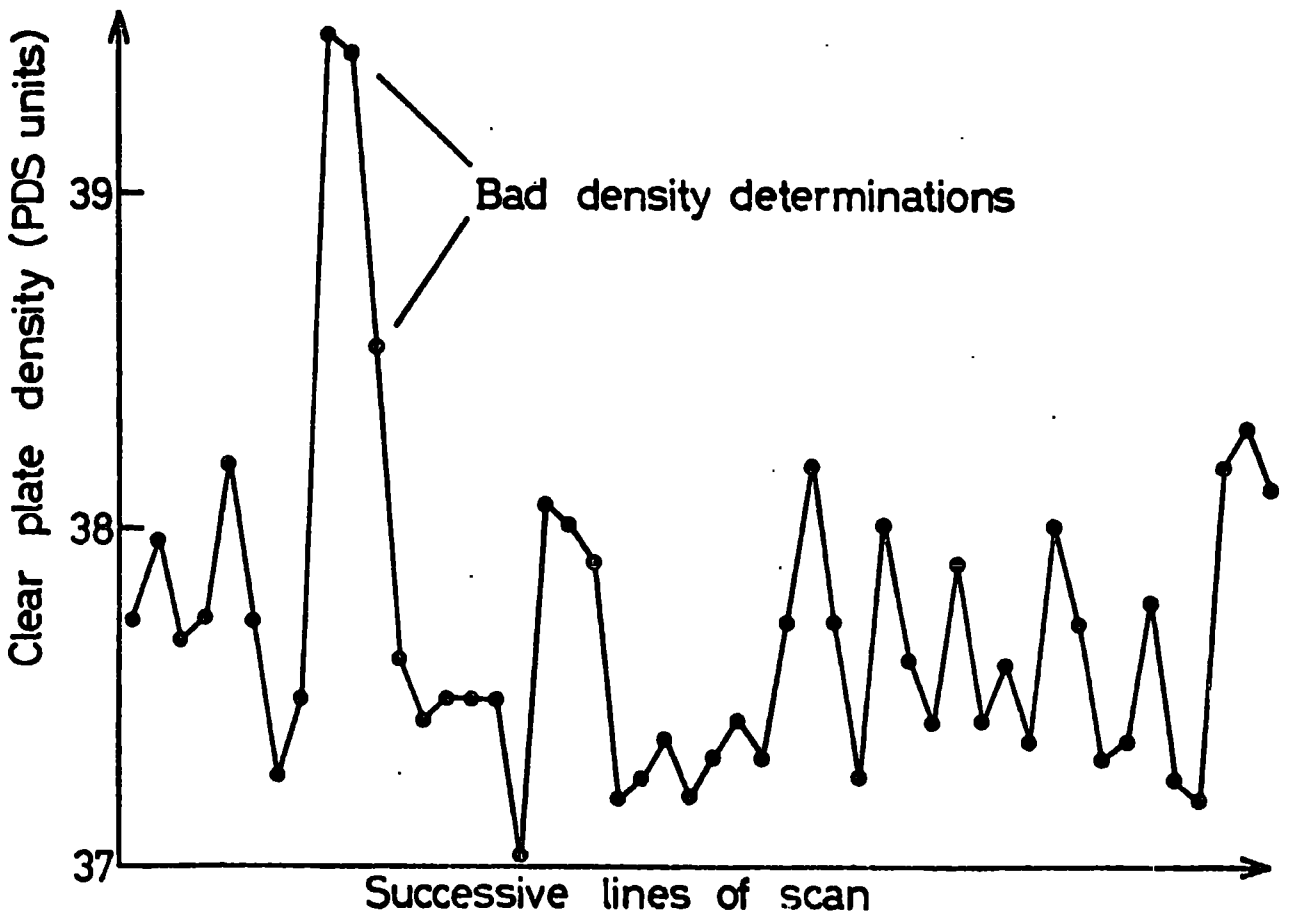


Fig. 5.3 Typical 'raw' clear plate trace

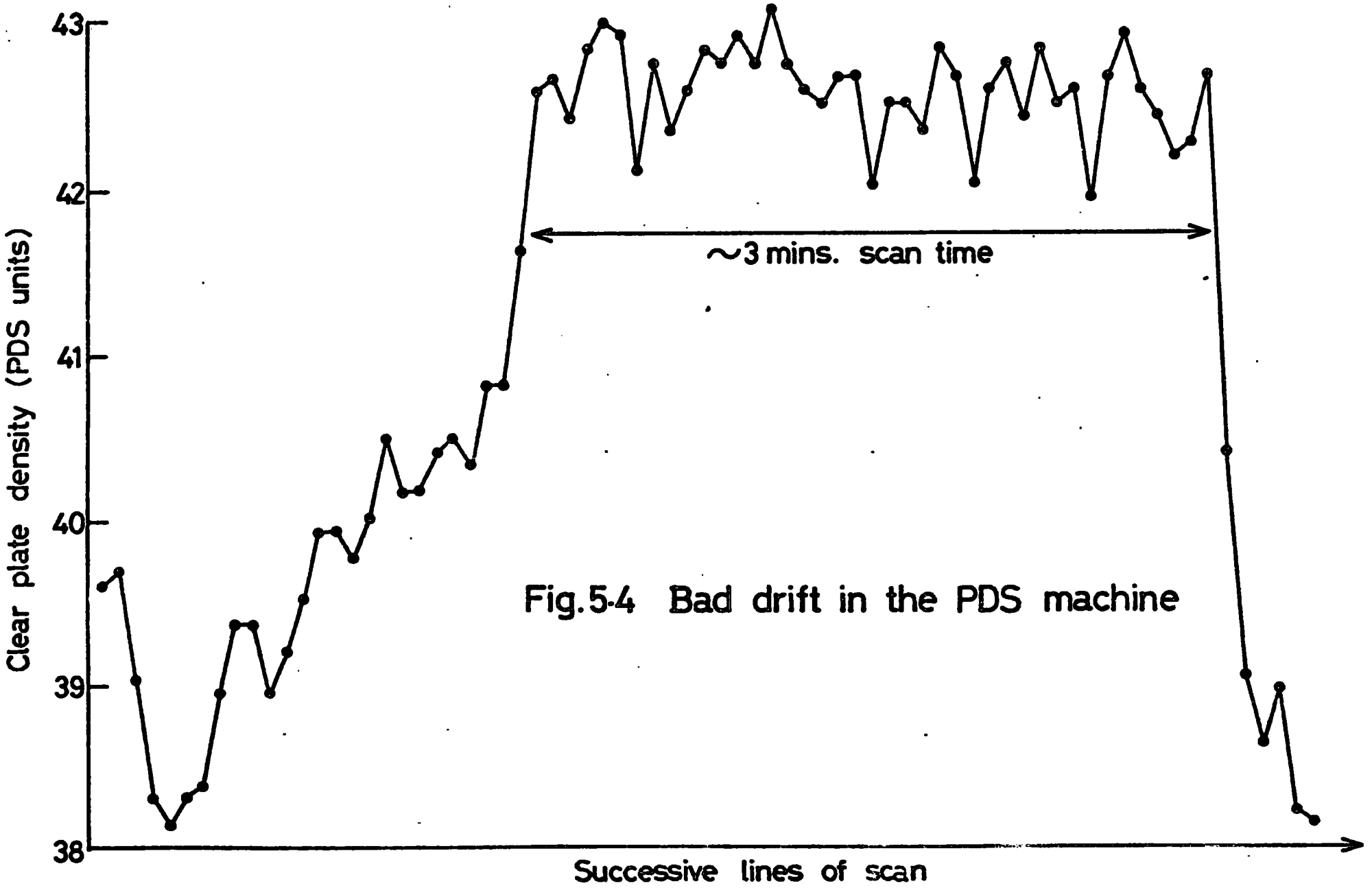


Fig.5-4 Bad drift in the PDS machine

sampled at the end of each line, the statistical error on the clear plate density calculated is quite high, and if a simple linear interpolation between the two ends of a line were employed, this would significantly increase the noise on the final result. What is required is a method of smoothing the clear plate trace over several lines to remove the statistical errors, while at the same time allowing the PDS jumps to remain and thus to be subtracted. In addition, the erroneous lines must not be allowed to join in the smoothing process or their effect will simply be spread to adjacent lines. The solution to this problem lies in a non-linear smoothing technique outlined below.

5.1.3 Processing the Clear Plate Traces

The first problem is how to apply local smoothing to the clear plate trace, while allowing abrupt jumps to remain. The solution adopted involves replacing each value by a weighted mean of its neighbours, the weights, however, are allowed to vary depending on the magnitudes of the jumps encountered between adjacent data points. Thus, if a small jump is encountered, this indicates no data corruption and a large weight is employed to provide good smoothing. A large jump, however, results in a small weight so that there is little smoothing and the jump is preserved. The weights are calculated in a sequential fashion, each being derived from the previous

one as successively more distant neighbours are considered, thus a large jump in the data results in all the neighbours beyond the jump being given small weights.

Mathematically, to replace a data point d_j, a series of weights w_k are calculated:

$$w_k = \prod_{i=1}^{i=k} \exp \left\{ \frac{-(d_{i+j-1} - d_{i+j})^2}{2(\sigma^2_{i+j-1} + \sigma^2_{i+j})} \right\} \qquad 5.1$$

for k = 1 to 7, and a similar set for negative values of k in the range -1 to -7. Here, at each point, the jump encountered is compared with that expected from statistical errors,

namely $(\sigma^2_{i+j-1} + \sigma^2_{i+j})^{\frac{1}{2}}$, and the accumulated weight is adjusted accordingly. The replacement value d'_j for d_j is then calculated as a weighted mean:

$$d'_j = \frac{\sum_{k=-7}^{k=+7} W_k d_{j+k}}{\sum_{k=-7}^{k=+7} W_k} \qquad 5.2$$

where $w_o = 1$. The estimates σ_j of the errors on d_j are processed using these weights to derive new estimates σ'_j of the errors on the replacement values d'_j. The effect of this non-linear smoothing is shown in figure 5.5 (b).

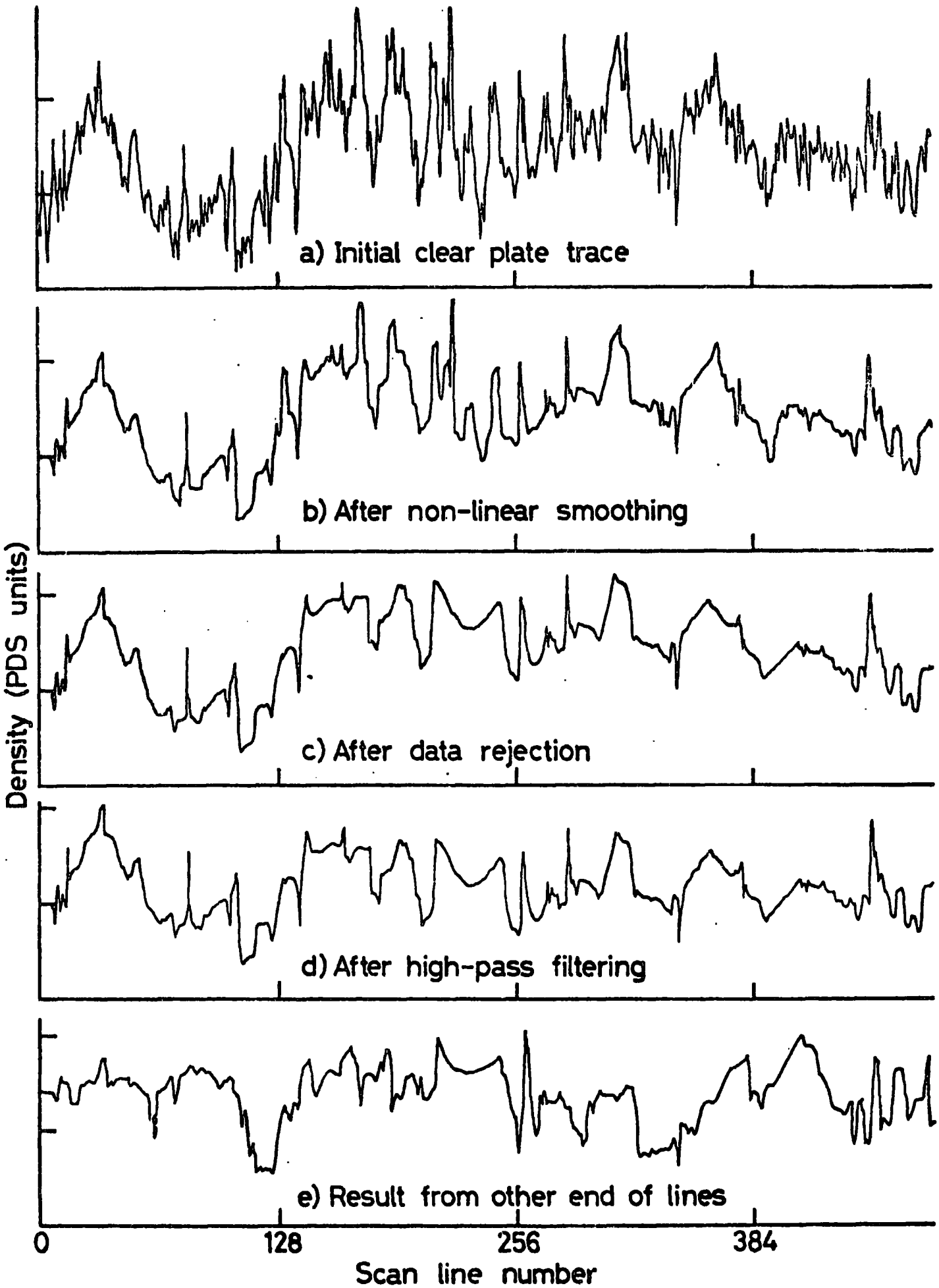


Fig. 5.5 Stages in processing the clear plate traces

The data now consists of a smooth background representing the clear plate density variations, with jumps superimposed due to the PDS machine and corrupt data. To distinguish between these jumps we use the fact that those due to the PDS machine affect both ends of a line, while the corrupt data is due to localised blemishes on the electronograph and affects only one end of a line. Thus if we subtract the values at opposite ends of each line, we get 512 values of a difference signal D_j, which should be smooth except for the corrupt data. This can now be detected and removed using the following procedure which is empirical, but gives good results.

A least-squares 5th. order polynomial fit is made to the difference signal by finding a polynomial $P_5(j)$ which minimises the function:

$$f = \sqrt{\left\{ \sum_j \left[\frac{(P_5(j) - D_j)^2}{(\sigma'^2_{jL} + \sigma'^2_{jR})} \right] \div (N - 6) \right\}} \qquad 5.3$$

where the subscripts L and R refer to the left and right hand ends of a line, and the summation is made over all N points which have not been rejected (See below).

To detect the jumps, we compare each value of the difference signal with the fit and reject it if:

$$| D_j - P_5(j) | > gf \qquad 5.4$$

where g initially has the value 1.8. Since we do not know which end of the line is in error, we reject both ends and replace them with estimates derived from the neighbouring lines (see next paragraph). The process is repeated with subsequent values of g being 1.6, 1.5, 1.5, 1.4, making the acceptance criterion progressively more strict, and is terminated after 4 repeats, or when f exceeds unity by no more than 10%, indicating that the remaining deviations from the fitted polynomial are approximately those expected on statistical grounds.

When data points are rejected in any of the above iterations, they are replaced, before proceeding to the next iteration, by a weighted mean of those neighbours which have not been rejected so far, and the difference signal D is re-calculated. In subsequent iterations, these points do not contribute to the fit (Equation 5.3), but they are candidates for repeated rejection and replacement (Equation 5.4). The weights used when replacing rejected values are derived in the same way as for the non-linear smoothing technique described earlier (Equation 5.1), except that there are now "holes" in the data where points have been rejected. This is dealt with by propagating the weights across the holes as though the intervening holes did not exist. Since it is not now

known how many neighbours to consider, sufficient are used for the weight of the most distant to have fallen to a negligibly small value.

The nett result of this rejection procedure is that points which have deviations from their surroundings greater than would be expected on statistical grounds are rejected, but so too are those which have unusually large statistical error estimates associated with them. This result is as required, since it is reasonable to expect that the statistical error estimates on each line should be approximately equal, and that a high value indicates some form of data corruption. The overall result is a clear plate trace which preserves the PDS drift and clear plate irregularities but is otherwise smooth (Figure 5.5c). There remains one effect, however, which needs further correction.

5.1.4 Image Proximity

When a clean clear plate trace is examined it often presents a rise in density at the centre relative to the two ends. Inspection of the raw data reveals that this is due to a small rise in density of the clear plate in the vicinity of the image on the electronograph; the centre of the trace being closest to the image, it thus shows this rise. The effect is small, typically ~1 PDS unit, but is larger when a particularly dense image is present.

The origin of this effect is not obvious, since it is often highly asymmetric about the central image and the asymmetry is neither constant, nor correlated with the settings of the polarimeter optics. It thus seems unlikely that it is due entirely to scattering effects in the polarimeter or electronographic camera. One possibility is that it arises during the developing procedure, since the electronographs are placed in the developing bath vertically - thus introducing a possible asymmetry in solution strength due to convective mixing in the image region.

Because this effect cannot be examined until the data - reduction stage it has not been fully investigated, indeed most users of the electronographic camera may be unaware of its presence. Whatever the cause, however, there is little doubt that it is associated with the image and should not appear on a "clear plate" trace. The method of removing this effect is to pass the trace through a "high pass" filter, as follows.

5.1.5 High Pass Filtering

To prevent any linear trend in the data being removed by this process, a straight line is fitted between the ends of the clear plate trace and subtracted from the entire trace. A heavily smoothed version of the result is then produced by replacing each value with a weighted

mean of all its neighbours. The weights follow an exponential law; to replace a data point d_j, we calculate weights w_k:

$$w_k = \exp\left(\frac{-|j-k|}{50}\right) \qquad 5.5.$$

and the replacement value d'_j is found from:

$$d'_j = \frac{\sum_k d_k w_k}{\sum_k w_k} \qquad 5.6$$

where the sum extends over all points in the trace.

The result of this smoothing is a trace which shows the unwanted rise in the centre, but which has all fine detail, including PDS jumps, smoothed out. When this is subtracted from the unsmoothed version, the unwanted effect is removed leaving the fine detail. Finally the linear trend on the data is added back to give the final clear plate trace (Figure 5.5d).

As can be seen from figure 5.5e which shows the final result for the opposite side of the same PDS scan, there is a large correlation between the irregularities on each side, indicating that they are due primarily to instability in the measuring machine. It is thought that with reasonably clean data, the clear plate background can be determined to within 0.5 PDS units by this

method, and it is hoped that by using the clear plate density as a monitor of PDS stability we have avoided the problems of non-reproducible measurements experienced by other users of the microdensitometer.

5.1.6 Linearity Corrections and Background Subtraction

Having determined the clear plate density at the end of each line of raw data from the PDS machine, the background can now be subtracted using linear interpolation along each line. At the same time, however, it is convenient to make corrections for the known non-linearities in the system.

Calibration of the PDS machine's response to density is achieved by measurement of a standard wedge of known density (the procedure is described in the PDS user's book at the Royal Greenwich Observatory) and a cubic polynomial has been fitted to the results. The form used is:

$$d_T = -0.4194 + 0.8966d_m + 0.00402d_m^2 + 0.0000003d_m^3 \qquad 5.7$$

where d_m is the measured density and d_T the true density. This correction is applied to both the clear plate traces and the data, and the background is then subtracted.

Finally, correction is made for the non-linear

response of the emulsion using the theoretical result of equation 4.30.

Before re-storing the data in integer form, it is scaled by multiplying by a factor 5 to prevent systematic rounding errors in subsequent stages of processing.

5.2 Correction for Photocathode Response

5.2.1 Introduction

At this point it is necessary to describe the method of making corrections for the spatial response of the photocathode of the electronographic camera without entering into details of how the response is initially determined, since this would require an explanation of later stages of the data reduction and thus would deviate from the logical path through the system. A full description of the photocathode calibration is given in a later section (5.5).

5.2.2 The Photocathode File

The photocathode response is stored in a computer file on magnetic disc known as the "photocathode file". Two such files are produced for each wavelength range used on an observing run, corresponding to the polarimeter "grids IN" and "grids OUT" settings, as these produce images on slightly different parts of the photocathode.

The file contains a 384 x 384 matrix of integers,

corresponding to the central image region of a 512 x 512 PDS scan and each integer represents the response of the photocathode at its particular location. This response is independent of the settings of the polarimeter since the light reaching the photocathode always has the same state of polarisation, and the response is thus constant throughout an observing run.

In addition to the photocathode response the matrix contains values which act as flags to indicate the absence of data at some points. These points include the region outside the polarimeter image where there is unexposed film, the regions where the images of the grid assembly overlap or fail to meet (the grid overlap regions) and any defects which can be identified on the photocathode where the data is unreliable.

Other quantities are stored in the photocathode file relating to the dimensions of the image, so that the file acts as a "template" to which each electronograph can be matched to identify the various elements of the image. A full description will be given when the photocathode calibration is discussed (5.5).

5.2.3 Alignment of the Photocathode Response

In order to correct with the photocathode response we must align the image to be corrected with the matrix in the photocathode file. This is achieved by using the "grid spots" as fiducial marks. The accurate positions

of these spots on the photocathode matrix are stored in the photocathode file, and the positions on the image to be corrected are found as follows.

5.2.4 Finding the Grid Spots

An initial estimate of the positions of the grid spots is made using the values for the photocathode matrix, since the images are aligned to within a few pixels at the time of digitisation. If this is not the case, corrections can be made to re-locate the initial estimates.

A 15 x 15 pixel area of the data matrix around the initial estimate is then taken, and X and Y density profiles of this area are formed by summing rows and columns within this area (Figure 5.6). The minimum values of these profiles are found and these backgrounds are subtracted from them. Their centroids are then found in X and Y and this gives an improved estimate of the grid spot location. The process is performed 3 times,each initial estimate being the centroid found on the previous iteration, and by this time the grid spot location is found to sufficient accuracy.

This technique will reliably locate grid spots even when the initial estimate is in error by as much as 8 pixels. Unfortunately, it will also locate regions of dirt so that each position must be verified before it is used.

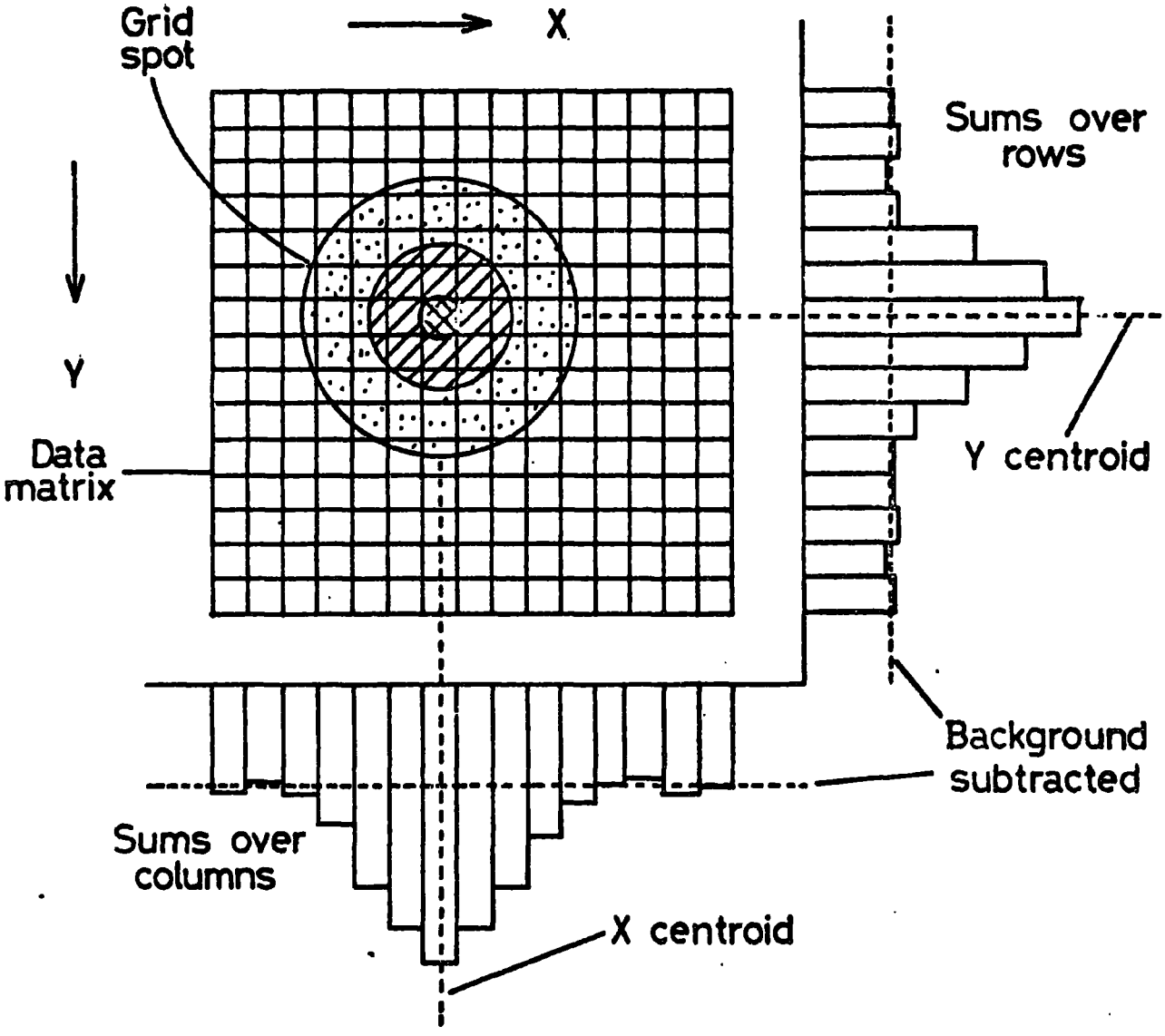


Fig.5.6 Locating the grid spots

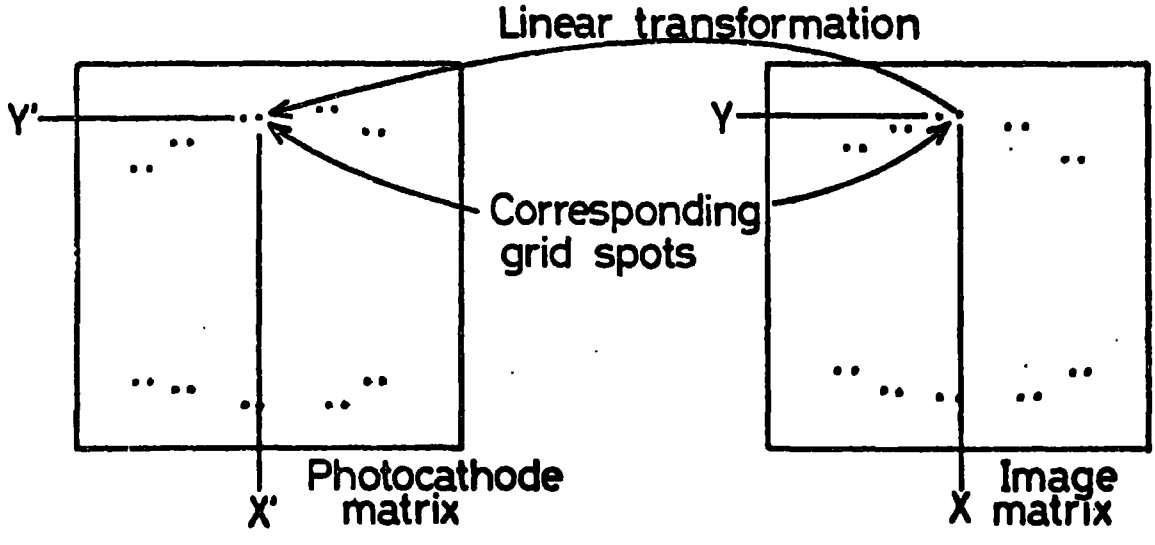


Fig.5.7 Relating positions on 2 images by a linear transformation

5.2.5 Determining the Transformation

Image alignment is achieved by finding a linear transformation which will convert an (X,Y) location on the image to the corresponding (X′, Y′) location on the photocathode matrix (Figure 5.7). The transformation employed has 6 free parameters (t_i, i = 1,6):

$$X' = t_1 + t_2X + t_3Y \qquad 5.8$$

$$Y' = t_4 + t_5X + t_6Y \qquad 5.9$$

which allows for a rotation and translation of the image and also for stretching in any direction as may be necessary if the electronographs change their dimensions with e.g. humidity or temperature.

A least-squares criterion is employed to determine the transformation. From the positions (X_i, Y_i) of the grid spots on the image, and those on the photocathode matrix (X_i', Y_i'), the transformation is found which minimises:

$$f = \sum_i \left[X_i' - (t_1 + t_2X + t_3Y) \right]^2 + \sum_i \left[Y_i' - (t_4 + t_5X + t_6Y) \right]^2 \qquad 5.10$$

The sum includes 20 grid spot images (except for any which are rejected - see below).

To check on the validity of the grid spot positions, the degree of mis-alignment δ is calculated for each:

$$\delta_i = \sqrt{\left[X_i' - (t_1 + t_2 X + t_3 Y)\right]^2 + \left[Y_i' - (t_4 + t_5 X + t_6 Y)\right]^2} \qquad 5.11$$

Since one bad position can cause the whole transformation to be in error, thus increasing the degree of misalignment for all the grid spots, only one position is tested for validity at a time by finding the largest value of δ, δ_{max}, i.e. the worst-aligned grid spot. This value δ_{max} is compared with the value 0.2 pixels, which is known to be an upper limit to the degree of misalignment found on good data, and the position is rejected if δ_{max} exceeds this. A new transformation is found, omitting this position, and the process repeated until the largest value of δ found is less than 0.2 pixels.

In practice 90% of the positions are usually accepted, and the RMS misalignment is typically 0.05-0.10 pixels, corresponding to an alignment accuracy of better than 2.5 μm.

5.2.6 Transforming the Photocathode Matrix

Having found the required transformation, we can now transform each position in the image to the corresponding location in the photocathode matrix. In general,

the transformed position will not lie on one of the grid points on which the photocathode response is known, so some form of interpolation between these points is required to determine the photocathode response to be used (Figure 5.8).

A discussion of the methods of performing this interpolation will be given when the data transformation is considered (5.3.8). In the present context a simple linear interpolation between the neighbouring 4 pixels is appropriate (Figure 5.9). This may be considered as a weighted mean of the neighbouring values, the weights being determined by the areas of overlap between the known pixels and the "required" pixel.

Where points on the photocathode matrix are flagged to indicate a region of no data, a decision must be made as to whether to use the corresponding point in the image. Figure 5.10 illustrates the criterion used; if the transformed position (X', Y') lies within a "bad" pixel, the corresponding pixel in the image is not used. If, however, it merely borders a bad pixel, the value is still used, the photocathode response being calculated by omitting the bad pixel (or pixels) and calculating the weighted mean of the reduced number of pixels. In this way, the number of "bad" pixels in the corrected image corresponds to the number in the photocathode matrix, and "bad regions" are not spread at the edges

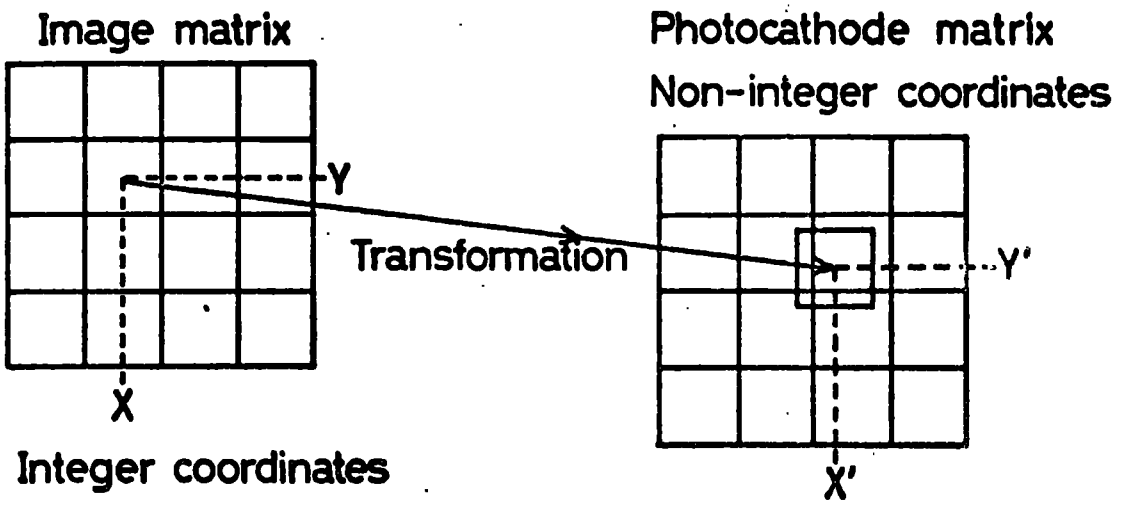
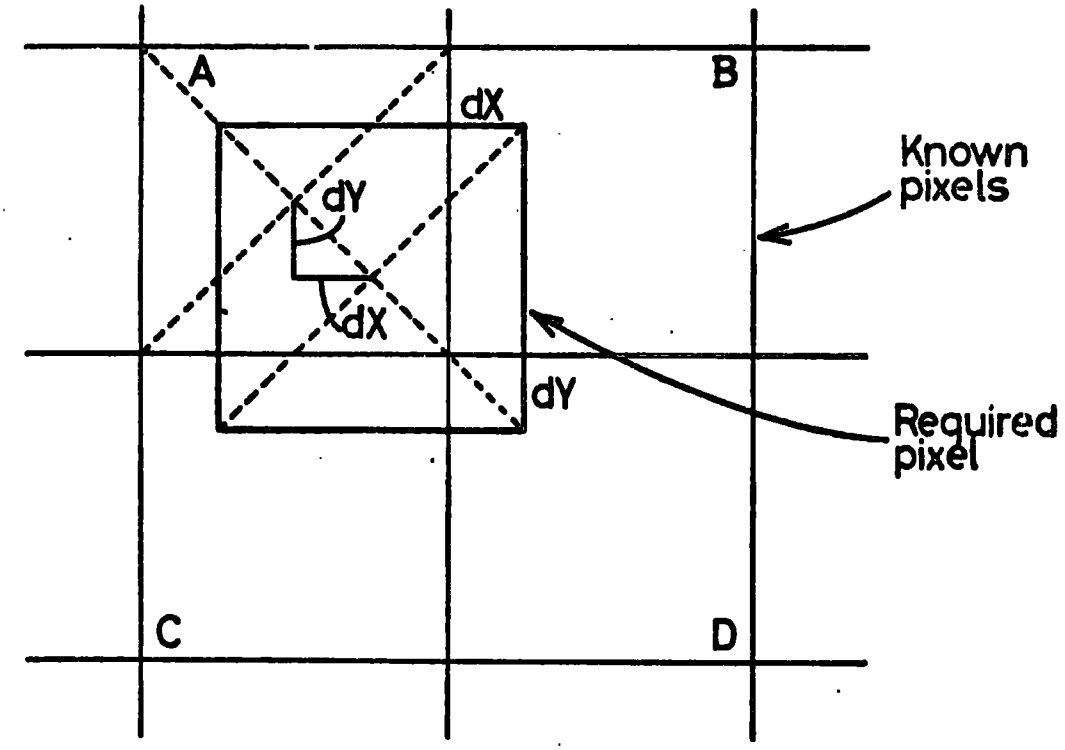


Fig.5.8 The need for image interpolation



Interpolated value = $(1-dX)\cdot(A\cdot(1-dY)+C\cdot dY)$
$+dX\cdot(B\cdot(1-dy)+D\cdot dY)$

Fig.5.9 Simple linear interpolation

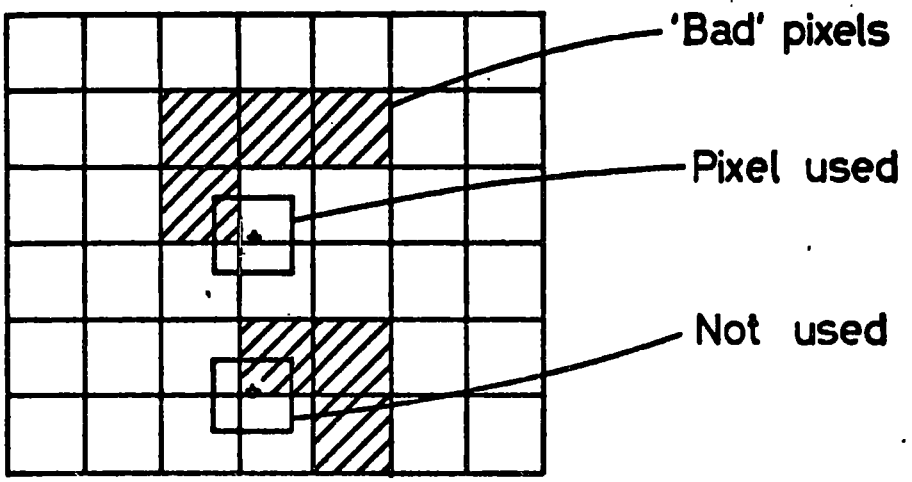


Fig.5·10 The criterion for rejecting transformed pixels

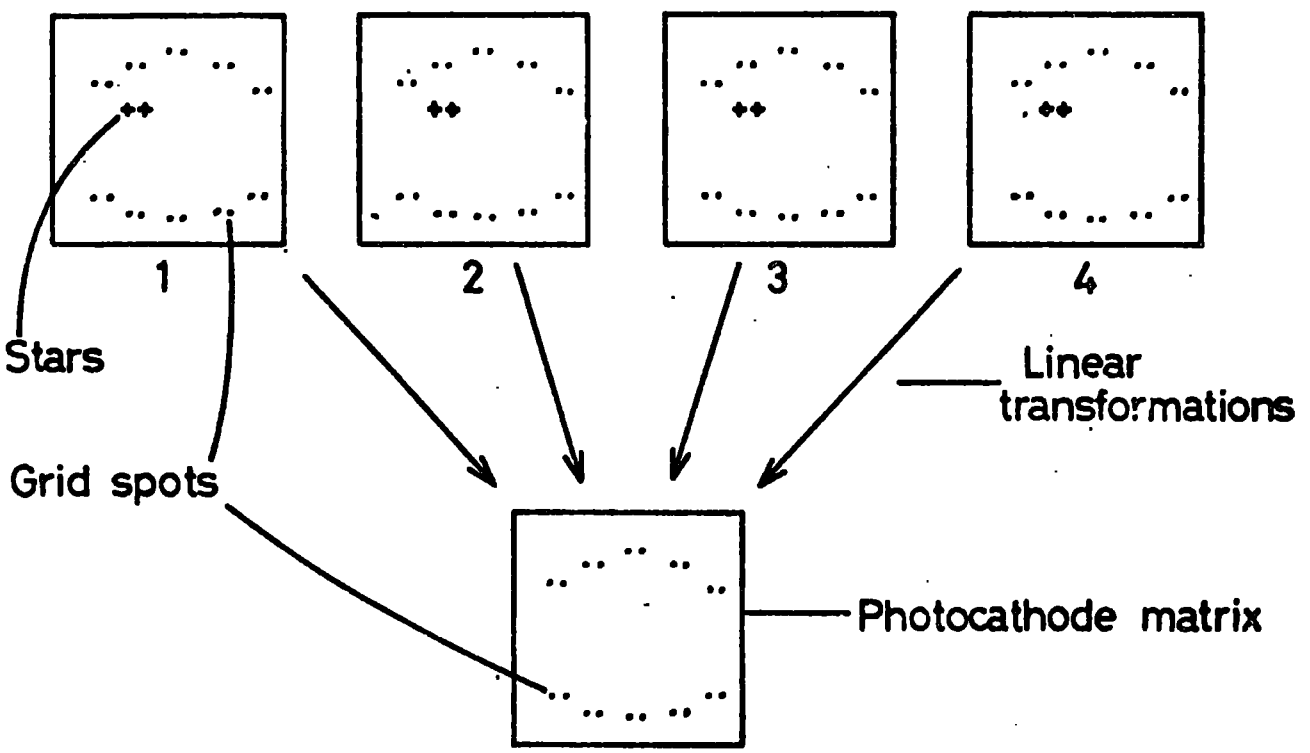


Fig.5·11 Aligning 4 images by aligning each with the photocathode reference frame

by the interpolation process.

5.2.7 Flagging Bad Data and Making the Photocathode Correction

The pixels in the image which are not used as a result of the above labelling of the photocathode matrix are flagged as invalid, and this flagging is carried through the rest of the reduction procedure. In this way, bad data detected at one stage is prevented from affecting later stages.

Before the photocathode correction is made, pixels which had an original reading on the PDS machine greater than 1000 are also flagged as invalid, thus removing regions of the image which are over-exposed. The photocathode correction is then made by dividing each valid pixel by the corresponding photocathode response.

5.3 Image Alignment

5.3.1 Introduction

Having recorded 8 images of an object in various polarisation states on 4 electronographs, we must ensure that we use identical regions of these images when we calculate the polarisation. Consequently we must align the images before we start to compare them with each other.

5.3.2 Fiducial Marks

We have already developed the technique of aligning two images using the grid spots while performing the

photocathode correction. It is a relatively simple matter to perform this procedure in reverse and transform an image to align it with the photocathode matrix (instead of vice-versa as was done previously). If we transform each of our electronographs to the photocathode matrix in this way, all 4 will then be aligned with each other. Thus we will have achieved the first step in aligning the 8 images (Figure 5.11). We then merely need to be able to transform the right hand image on each electronograph to align it with the left hand image.

Unfortunately, such a simple process is not sufficiently accurate for our purposes. The difficulty arises because, in aligning the grid spots on separate electronographs, we are merely aligning the separate images of the polarimeter grid assembly. This does not necessarily mean that the separate images of the object under study will be aligned, because of the possibility that the object has moved in the polarimeter field between successive electronographs. Such movement could result from imperfect guiding of the telescope or from flexure, probably in the polarimeter mounting on the telescope.

To overcome this problem, we need a set of fiducial marks within the object itself, so we use the images of stars within the polarimeter field. While the separate

electronographs could be aligned with each other using the stars alone (e.g. Axon 1977, Pallister 1976), it often happens that the number or positioning of the stars within the field is far from ideal for calculating a set of 6 transformation coefficients. Consequently, a technique is employed which uses both grid spots and stars simultaneously, while allowing the stars to move relative to the grid spots. In this way, the rotation and stretching part of the transformation is defined mainly by the grid spots, and the dependence on the stars is greatly reduced.

5.3.3 The Reference Frame

Rather than align 3 of the electronographs with the remaining one, as has been done in a previous reduction technique (Axon 1977), there are advantages to aligning all 4 electronographs with a neutral reference frame i.e. the photocathode matrix. The obvious advantage of this is that the photocathode matrix can initially be chosen so that the polarimeter image is in a known location, with the grid bars running parallel to the columns of the matrix (the Y axis). Any electronograph aligned with this matrix will thus also be in a similar position within its matrix, so that the various components of the image can be immediately located.

A less obvious advantage is that all 4 electronographs

undergo the same number of transformations (namely 1). There is necessarily some data degradation in the transformation process and this is thus made equal on each image, an important consideration for the later stages of data validation and error calculation.

5.3.4 The Alignment Procedure

We are concerned initially only with finding the coefficients of the linear transformations required to align the images. Only when all the transformations are known and have been combined into a simple set of 8 transformations, one for each image, are the images actually aligned.

The first step in the alignment procedure is to find a set of star images on each electronograph. These are easily located on computer line-printer pictures of the data (e.g. Figure 3.3) and the accurate positions are found using the same technique as for locating the grid spots which are a similar size on the electronographs (5.2.4).

5.3.5. Aligning Plate 1

The first electronograph to be considered (referred to as plate 1) does not have to align with any stars, and it is simply aligned to the photocathode matrix using the grid spots alone. The transformation required is simply the inverse of that used for the photocathode

correction (5.2.5).

To provide a reference set of star positions for the alignment of subsequent electronographs the positions of the stars on the first electronograph are transformed to give the positions they will occupy in the photocathode frame when the electronograph is aligned with the photocathode matrix.

5.3.6 Aligning Plates 2,3 and 4

Subsequent electronographs must be aligned both with the reference set of grid spots in the photocathode matrix and with the reference set of star positions in the photocathode frame which have been derived from the first electronograph.

The alignment procedure used involves the linear transformation used previously, but with 8 free parameters ($t_i, i = 1,6$), $\Delta x, \Delta y$ relating a position (X', Y') in the photocathode frame of reference to a position (X,Y) on one of the electronographs to be aligned:

$$X = \Delta x + t_1 + t_2 X' + t_3 Y' \qquad 5.12$$

$$Y = \Delta y + t_4 + t_5 X' + t_6 Y' \qquad 5.13$$

The extra coefficients $\Delta x, \Delta y$ represent the movement of the star positions relative to the grid spots (i.e. relative to the polarimeter field) since the first electronograph was taken.

A more involved least squares criterion is used for determining this transformation, namely minimisation of:

$$f = \sum_i \left[X_i - (t_1+t_2X'_i+t_3Y'_i)\right]^2 + \sum_i \left[Y_i - (t_4+t_5X'_i+t_6Y'_i)\right]^2$$

$$+ \sum_j \left[x_j - (\Delta x+t_1+t_2x'_j+t_3y'_j)\right]^2 + \sum_j \left[y_j - (\Delta y+t_4+t_5x'_j+t_6y'_j)\right]^2$$

5.14

where (X_i, Y_i) and (x_j, y_j) are the positions of the grid spots and stars on the electronograph to be aligned and the dashed quantities are the corresponding positions in the photocathode reference frame. The first 2 terms are summed over those grid spot positions which are known to be valid from the previous photocathode correction stage and they represent the usual 6 parameter transformation. The last 2 terms, which include the shifts in X and Y location $\Delta x, \Delta y$ are summed over the star positions. At this stage it is not known if any of the star positions are invalid, so a rejection procedure for the star positions is adopted as follows.

The RMS misalignment of the N stars is calculated as:

$$\langle \delta \rangle = \sqrt{\frac{1}{N-1}\left[\sum_j \left\{x_j - (\Delta x + t_1 + t_2x'_j + t_3y'_j)\right\}^2 + \sum_j \left\{y_j - (\Delta y + t_4 + t_5x'_j + t_6y'_j)\right\}^2\right]}$$

5.15

and those stars which have a misalignment in X position, or in Y position of more than $1.8\langle\delta\rangle$ are rejected. The transformation is re-calculated and the process is repeated twice more, omitting the rejected stars each time. At the end of this rejection process, the 8 parameter transformation is converted to a 6 parameter one by combining the constant terms, resulting in the required transformation.

5.3.7 The Left-Right Transformation

As a result of the above procedure we can find a set of 4 transformations which relate a position (X',Y') on the photocathode matrix to the corresponding positions (X,Y) on each of the 4 electronographs, such that when the electronographs are transformed in this way their star images (but not necessarily grid spots) will co-incide.

The next step is to include the transformation between the left and right hand images on each electronograph so that all 8 images can be aligned. The information required is a transformation which relates a position on the left hand image on the photocathode matrix to the corresponding position on the right hand image. As figure 5.12 shows, when this is known, it is possible to calculate algebraically the transformation from the left hand position on the photocathode matrix to the right hand position on any of the images. Since we already know the transformation to the left hand position,

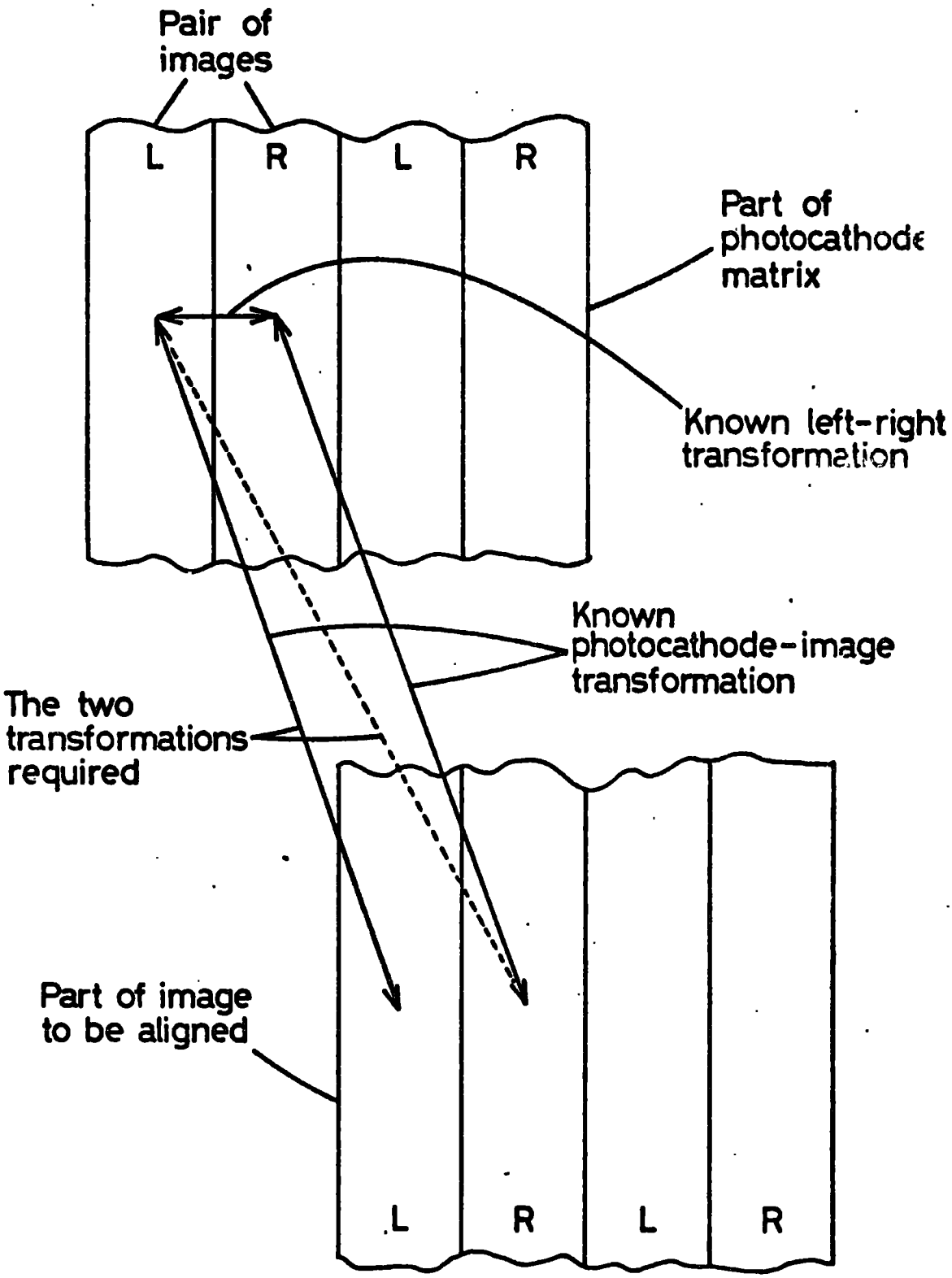


Fig.5.12 Including the L-R transformation to align the left and right hand images

we can now find the complete set of 8 transformations required to align all 8 images in the photocathode reference frame.

The left-right transformation required, because it is referred to the photocathode frame, is constant throughout an observing run since it depends only on the polarimeter alignment and the position of the image within the photocathode matrix. Hence it is stored in the photocathode file and called on when required. It is derived initially from the grid spots in the photocathode matrix and this process is described in the section on the photocathode calibration. (5.5)

5.3.8 Image Interpolation

Having found all the linear transformations required to align the images, the next step is that of actually transforming the data. The same problem arises here as in the photocathode correction, namely, positions in the reference frame which lie at the centre of a pixel do not lie at the centre of a pixel when transformed to give the corresponding position in another matrix. (Figure 5.8). We must therefore interpolate between pixels in order to effect the transformation. A number of methods of doing this have been suggested and all have their advantages and disadvantages. Some of these will be considered here.

The first method - deconvolution - has its basis

in the theory of sampling. It can be shown (e.g. Lynn, 1973, p160) that if a signal is "band limited" (i.e. its Fourier transform is zero outside some range) and it is sampled sufficiently frequently on a rectangular grid, then the intermediate values between grid points can be reconstructed with zero error. The frequency with which sampling must be carried out for this to be true is called the Nyquist frequency, and this depends on the range in which the Fourier transform is non-zero, or in practice on the fineness of detail in the signal. In effect, astronomical signals are band limited since the fine detail is "washed out" by the effects of atmospheric turbulence, so this method could be employed.

The problem lies in a practical realisation; to reconstruct the signal the sampled values must be convolved with a 2-D sinc function with appreciable extent, resulting in excessive computing time requirements. Furthermore, the presence of "edge effects", particularly since the polarimeter grid assembly segments the image, precludes its use here.

An alternative approach, suggested by Axon (1977), that of fitting 2-D cubic splines to local regions of the data, is potentially more useful since it is computationally more efficient and edge effects can be coped with.

Both the above methods, however, suffer from a serious drawback in the context of electronography - they both use a large region of data in calculating the

interpolated value, hence any corrupt data (which a cubic spline will not fit) is spread by the interpolation process into the surrounding area of data, making it extremely difficult to detect.

Since electronography suffers from a high incidence of corrupt data, the best method must use as small a region as possible in making the interpolation.

5.3.9 Linear Interpolation

The method which uses the smallest region of data is the linear interpolation between the 4 adjacent pixels, as used in making the photocathode correction, and it is unlikely, in the presence of noise, that it is significantly less accurate than the above methods, particularly since an average over several adjacent pixels is formed when the final polarisation map is produced.

This method would be ideal, were it not for the following drawback - if each pixel in the neighbourhood has an independent measurement error σ, then the probable error on the result is:

$$\sigma_R = \sigma\sqrt{\left[(1-dx)^2(1-dy)^2+dx^2(1-dy)^2+dy^2(1-dx)^2+dx^2dy^2\right]} \qquad 5.16$$

(notation as in figure 5.9). As can be seen, this depends on the values of dx and dy, which will be different in different parts of the image. The factor σ/σ_R is a measure of the information content of the result and

represents smoothing of the image. Hence, some parts of the image are smoothed, and other parts (where dx and dy are zero) are not smoothed at all. Since there is no way of later determining how mush smoothing has taken place, there is no way of estimating the errors on the interpolated values (on which the data validation stage relies), nor of estimating the errors in the final polarisation map.

5.3.10 The Constant Noise Interpolation

The solution adopted is to retain the linear interpolation but with a modification such that the probable error on the interpolated value is constant. This results in a constant amount of smoothing of the entire image and hence a constant known resolution across the entire field.

As figure 5.13 shows, the method is the same as the simple linear interpolation, except that the size of the "required" pixel is increased to cover more of the data and hence provide more smoothing. By adjusting the size of this pixel according to the values of dx and dy, the required constant noise on the result is obtained.

With the notation of figure 5.13, the error on the interpolated value is given by:

$$\sigma_R^2 = \frac{\sigma^2}{16\, \ell_1^2\, \ell_2^2} \; (2r_1^2+2dx^2+1)\; (2r_2^2+2dy^2+1) \qquad 5.17$$

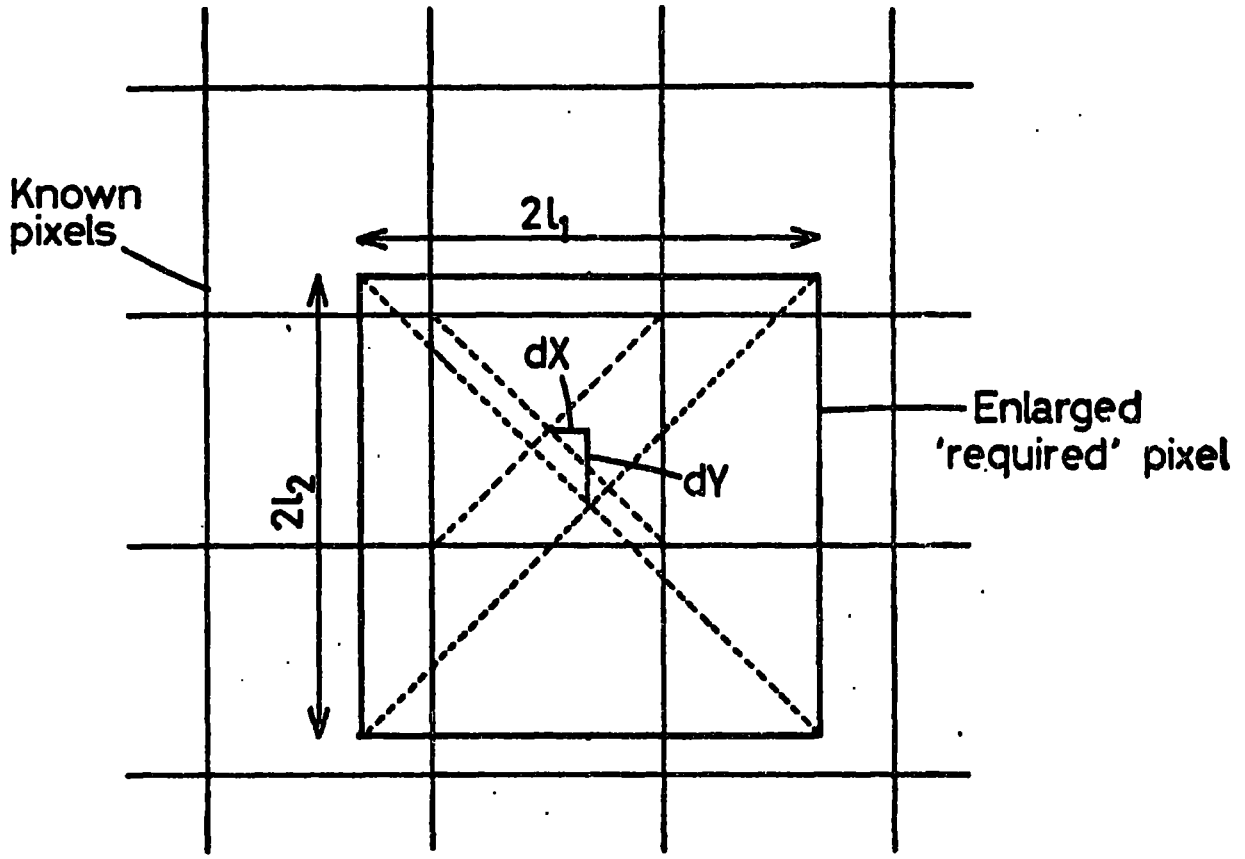


Fig.5.13 The constant noise interpolation

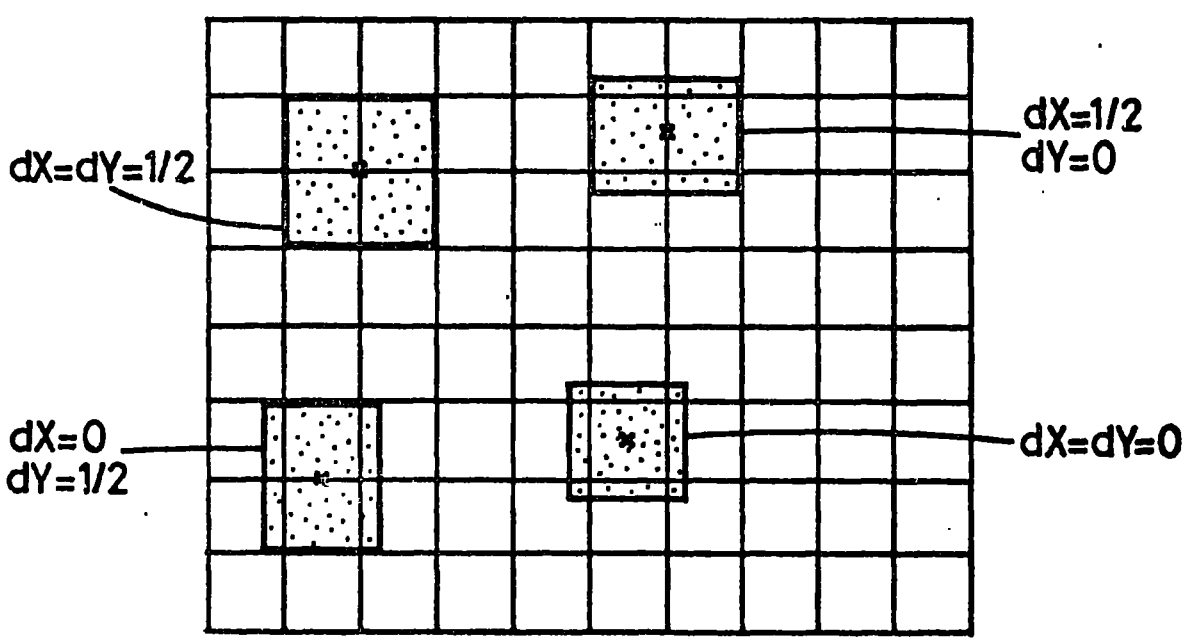


Fig.5.14 Special cases of the interpolation

where $r_1 = \ell_1 - \frac{1}{2}$ and $r_2 = \ell_2 - \frac{1}{2}$.

The maximum amount of smoothing given by the simple interpolation technique (Equation 5.16) gives $\sigma/\sigma_R = 2$, hence, since we can only provide increased smoothing to keep σ/σ_R constant, we aim to make $\sigma/\sigma_R = 2$ everywhere. Together with the need for symmetry in the X and Y directions, this requires:

$$\frac{2r_1^2 + 2dx^2 + 1}{4\ell_1^2} = \frac{2r_2^2 + 2dy^2 + 1}{4\ell_2^2} = \frac{1}{2} \qquad 5.18$$

which gives:

$$\ell_1 = dx^2 + \frac{3}{4} \qquad 5.19$$

$$\ell_2 = dy^2 + \frac{3}{4} \qquad 5.20$$

Thus, in general, the "required" pixel must be made rectangular. Some special cases of the interpolation are shown in figure 5.14, where the action can be seen. While a slightly larger area of data is used than with the simple interpolation, the extra pixels are included with only a small weight, and corrupt data is not significantly more spread.

The occurrence of pixels flagged as invalid is dealt with as in the case of the photocathode correction, i.e. if the transformed point lies within an invalid pixel, the corresponding point in the transformed image

is flagged as invalid, otherwise, neighbouring invalid pixels are simply omitted when forming the weighted mean (Figure 5.10).

5.3.11 Transformation and Storage of Data

Using the above interpolation technique and the transformation coefficients determined previously, the 8 individual images are transformed to the photocathode matrix frame of reference so that they are all aligned.

At this point it becomes necessary to adopt another method of storing the data, since it can no longer be contained in 4 square matrices. The method adopted is to reject all unwanted data (i.e. clear plate regions) and split the remaining data into 6 regions corresponding to the apertures in the grid assembly (Figure 2.2). These blocks of data (referred to as "grids" of data henceforth) then contain 8 images each, the values of all 8 images at any one point being stored in adjacent locations. This arrangement is efficient in the use of storage and processing time and the entire set of data is contained in one computer file. Because of the large amount of data in this file it is rarely possible to load the entire contents of the file into the computer store. A more convenient and more efficient way is thus adopted of processing 1 grid of data at a time in later stages of the reduction technique.

5.4 Sky Signal Subtraction

5.4.1 Introduction

At this stage, the data still contains a background signal, due to the light from the night sky, which must be subtracted to give the contribution from the object under study alone.

The accurate determination of the sky signal is of great importance, since often the regions of interest have brightnesses which are a fraction of the sky brightness and the results are thus greatly affected by the amount of background subtracted. Furthermore, an erroneous sky signal will affect the whole of one image, leading to an entirely erroneous polarisation map. Thus it is important not only to determine the background signal accurately, but to provide checks in later stages to make sure that this has been done successfully.

If the sky brightness is uniform across the polarimeter field, then having corrected for the photocathode response, the background signal should be constant across the entire image. A different background signal must be determined for each of the 8 images, however, because of the possible changing polarisation and brightness of the night sky.

5.4.2 Sampling the Sky Signal

No evidence has been found to suggest that the sky

brightness cannot be assumed constant across the polarimeter field to the required accuracy ($\sim \pm$ ½%), consequently we merely need to sample the sky signal at suitable locations in an image and subtract this value from the entire image. The locations to be sampled are specified by the user of the data reduction programmes (Figure 5.15).

Since the same regions are sampled on all 8 images, it is necessary to use the aligned images when sampling the sky signal. The alignment process, however, introduces some spreading of corrupt data, as discussed previously, so for optimum rejection of corrupt data the alignment process is slightly modified when the sky signals are being determined. The alignment here does not need to be accurate to better than 1 pixel, so the interpolation technique (5.3.10) is omitted, and the nearest pixel to the transformed position is used, thus eliminating any smoothing due to interpolation.

The selected pixels are subjected to the following rejection technique to remove corrupt data and any star images which may be contained in the sky area.

5.4.3 The Rejection Process

The mean and standard deviation of all the sampled pixels within a given sky region is calculated (typically over 1000 pixels are used), and those pixels with a value greater than 1.5 standard deviations from the mean are rejected. The mean and standard deviation are then re-calculated

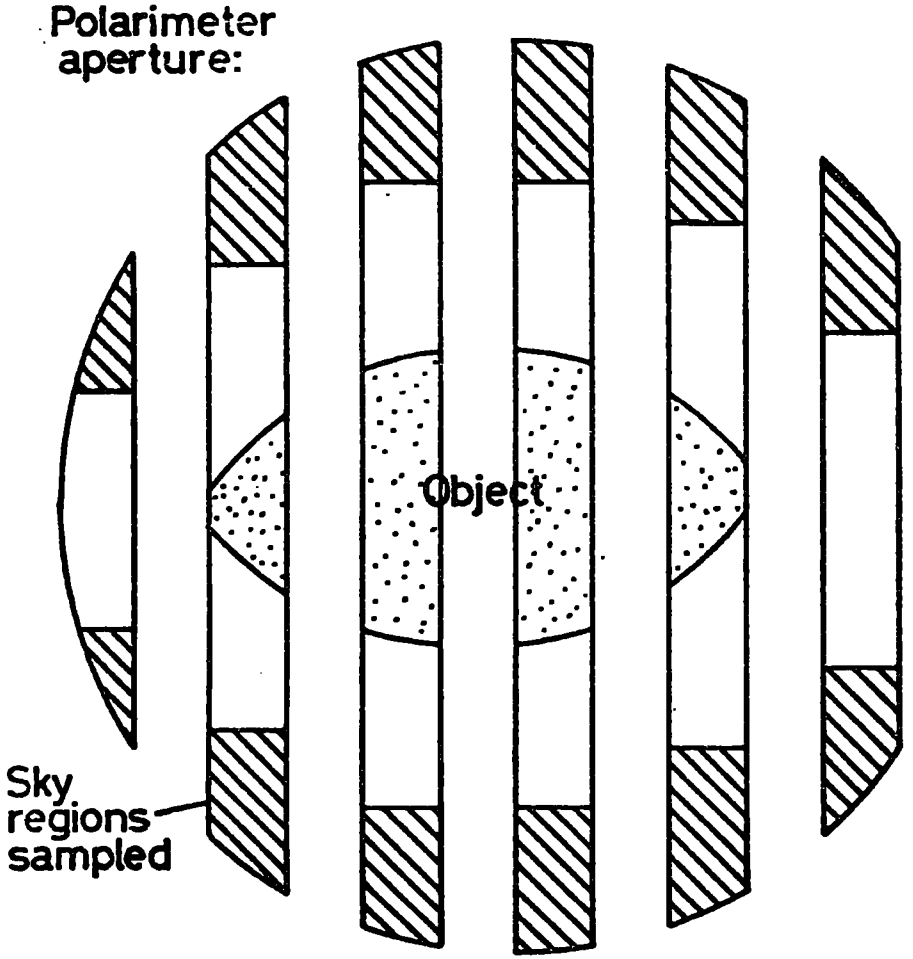


Fig.5.15 Typical sky regions sampled

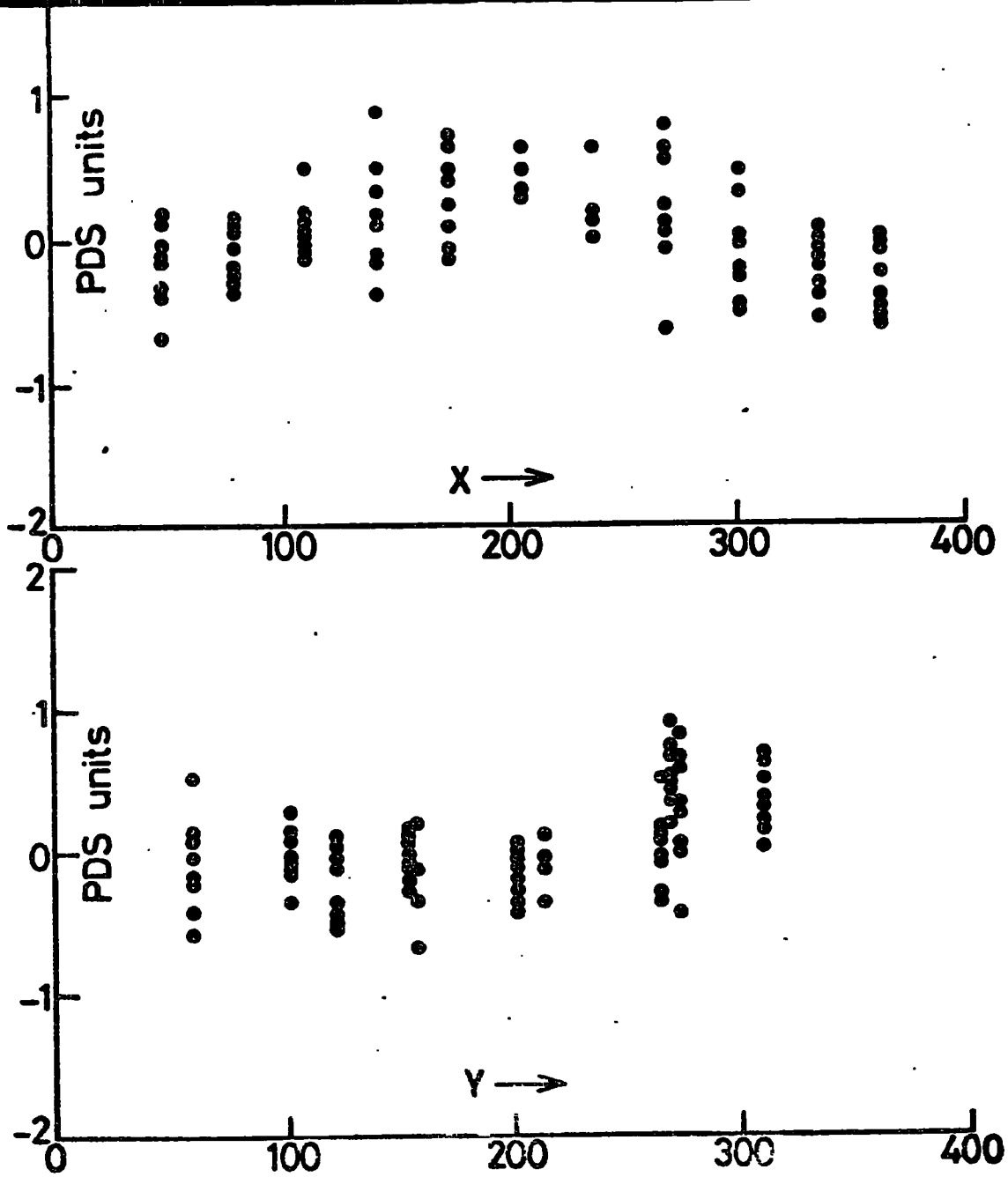


Fig.5.16 Typical sky residuals

omitting the rejected points.

The original, un-rejected set of pixels is then tested again, and those with values more than 1.5 new standard deviations from the new mean are rejected. The mean and standard deviation are re-calculated and the process repeated until either 10 successive means have been calculated, the same result has been obtained twice in a row, or more than 20% of the original pixels are rejected at any stage. (The advantage of testing the entire set of pixels at each stage is that initial errors in the mean, due for example, to a bright star in the area, do not have a permanent effect on later iterations.) The process is made deliberately strict since the major errors arise not from having a small number of data points (statistical errors), but from the corrupt data, which must be rejected even at the expense of good data.

This process is repeated for each region of the image specified by the user (Figure 5.15) and on each of the 8 images. In each region a mean value is found and an estimate of its statistical error made.

5.4.4 Testing the Different Sky Regions

The different regions sampled on any one image should have the same value of sky signal within some small error. To test this, the mean and standard deviation of the set of values obtained in these regions are calculated and

those regions with values more than 2 standard deviations from the mean are rejected. The mean and standard deviation are re-calculated and the original set of values is re-tested, as in the rejection process within the regions, except that only 3 iterations are used here.

Finally, the mean of all regions accepted is used as the sky signal and is subtracted from the appropriate image.

5.4.5 Sky Subtraction Errors

Usually, the values obtained in the different regions sampled have a dispersion somewhat greater than would be expected from the statistical errors within a region. To investigate this, the deviations of the sky regions from the mean adopted value (the sky residuals) are plotted against X and Y position in the image (Figure 5.16). This provides a check on the accuracy of the photocathode correction, as any error in this will be reflected as a systemmatic trend on these plots. Usually any systemmatic trend is of the same order as the intrinsic dispersion of the points and so is not considered significant. It is unlikely that the photocathode response is responsible for these trends, since they do not appear on all data corrected with the same photocathode response.

The magnitude of the sky subtraction errors appears to be given approximately by $\pm$0.5 PDS units $\pm$0.5% of sky signal. Since the errors are correlated on the two

images recorded together on one electronograph, however, the polarisation measurements, which depend on the difference between these 2 images, are somewhat more accurate than these figures would suggest.

5.5 Calibrating the Photocathode Response

5.5.1 Introduction

The description of the calibration of the photocathode response has been left until now since it involves the techniques developed for the reduction of the data which have been described in the earlier parts of this chapter.

5.5.2 The Basis of the Photocathode Calibration

As was described in chapter 2 (2.6.3), the spatial response of the photocathode of the electronographic camera is measured by taking electronographs using uniform illumination of the polarimeter field with light in the spectral range of the observations. These electronographs, after correction for the known errors in the system, provide a measure of the photocathode response at each point in the image and can be used for the correction of other data. There remains the problem, however, of how to present the response for this purpose; in particular, how to reduce the noise on these electronographs and remove the effects of blemishes so that data corrected with the response will not suffer as a result.

Other workers have used 2-dimensional polynomial fits to represent the photocathode response, this allowing the possibility of rejecting and smoothing over defects on the calibration electronographs and also ensuring smoothness and low noise level in the result. Practical experience of the technique, however, has shown that in order to achieve high accuracy, a large number of free parameters are required in the fit. (A 10th order fit - with 55 free parameters - was of debatable adequacy) High order fits, such as are required, involve much computing time and numerical problems arise because of the high accuracy required in computing the parameters of a polynomial fit.

The main reason for the difficulties encountered with polynomial fits is that there can be complex small-scale structure in the response which such fits cannot accommodate. Accordingly, some form of data-smoothing which has a small scale length is required. The obvious solution is a straightforward "moving average" type smoothing, since this can be adjusted to give the degree of smoothing required. In the context of polarimetry, however, such smoothing is superfluous, since the data which has been corrected with the photocathode response will eventually be smoothed by forming integrated measurements over large areas when the final results are calculated; thus smoothing the effective photocathode response at the same time. Hence, any further smoothing of the photocathode

response at the calibration stage can have only a small additional effect.

Since neither large nor small scale smoothing is appropriate as a means of reducing noise while maintaining an accurate description of the response, we must seek other methods of reducing the noise level and detecting blemishes on the electronographs. This can only be done by making intrinsically accurate measurements and the method employed is to combine a set of 4 electronographs, each of which has been exposed to a high density to achieve a good signal to noise ratio. At the same time, comparison between the 4 electronographs provides a means of detecting blemishes. While no smoothing of the data is employed to produce the final photocathode response, detection of blemishes is aided by requiring the response on each electronograph to be smooth on a small scale.

The details of the production of the response will be described in section 5.5.11, first we must consider the details of the alignment of the electronographs.

5.5.3 Alignment Considerations

At the photocathode calibration stage (the first step in the data reduction), we have no reference marks for image alignment. In particular, we do not know the positions of the grid spots on the electronographs, the positions of the features of the image, nor even if

the grid bars are running parallel to the Y axis as required. The first step, therefore, is to locate the vital features and produce a "template" describing the image. This will later form the basis of a photocathode file.

5.5.4 Locating the Grid Spots and Grid Corners

The probable positions of the grid spots on one of the 4 electronographs to be used for the photocathode calibration are found initially by looking at line-printer pictures of the data (e.g. Figure 3.3). The areas of interest are then inspected using larger scale line-printer pictures with a 1 pixel resolution and the grid spot positions located to within 1 or 2 pixels by eye (Figure 5.17). Subsequently, accurate positions are found using the technique of section 5.2.4 using the eye estimates as initial search positions.

On the remaining 3 electronographs, only 2 grid spots need be found by eye, as the accurate positions found on the first electronograph are then aligned to these 2 positions providing a complete set of 20 approximate positions so that all the grid spots can be found automatically. We then have complete sets of 20 accurate grid spot positions for each electronograph.

For purposes of aligning the grid bars parallel to the Y axis, each electronograph is inspected (using the 1 pixel resolution line-printer pictures) to locate

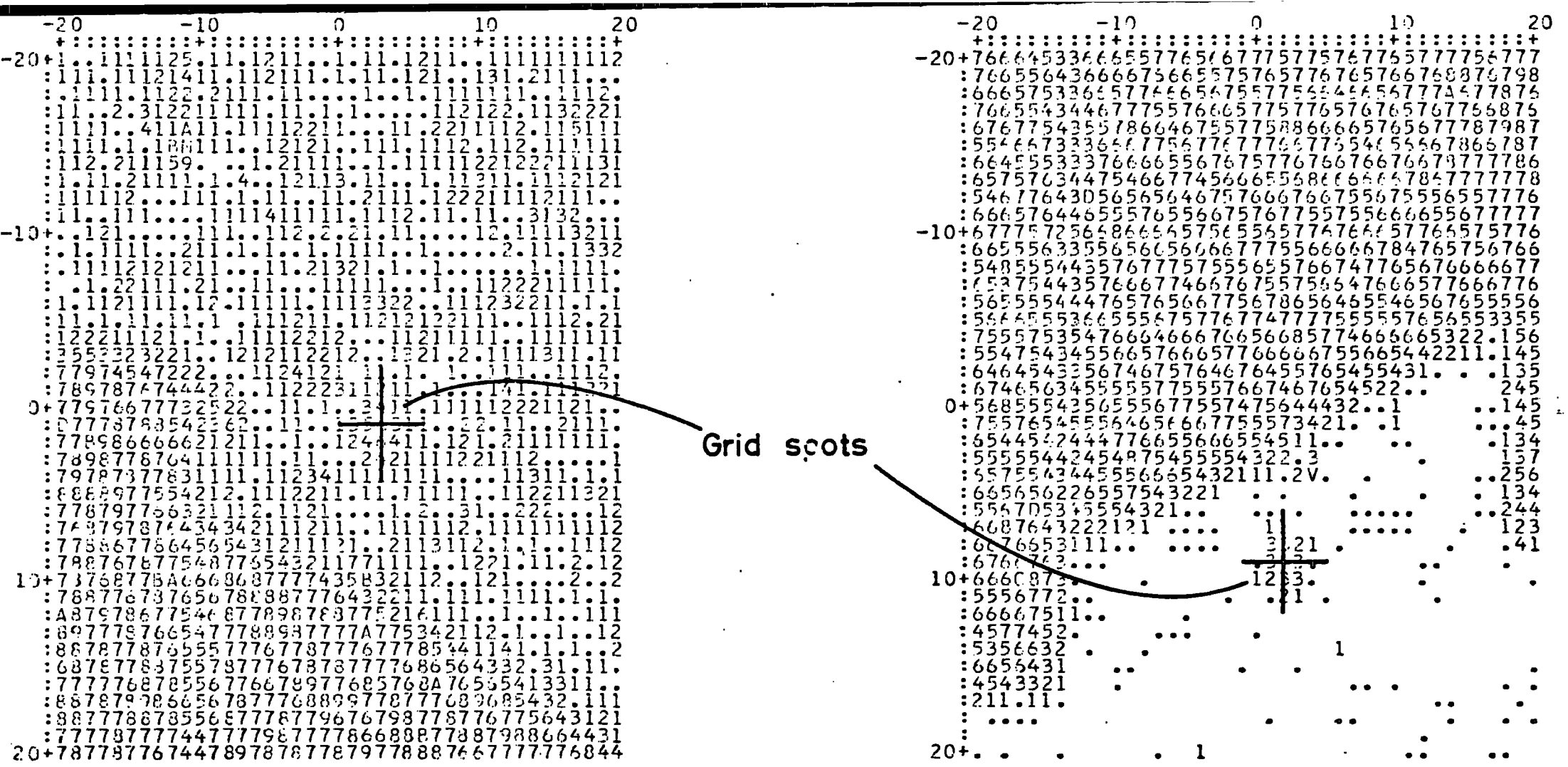


Fig.5-17 Locating grid spots by eye

the top and bottom corners of one grid aperture (Figure 5.18). This is done by eye to a probable accuracy of ~1 pixel.

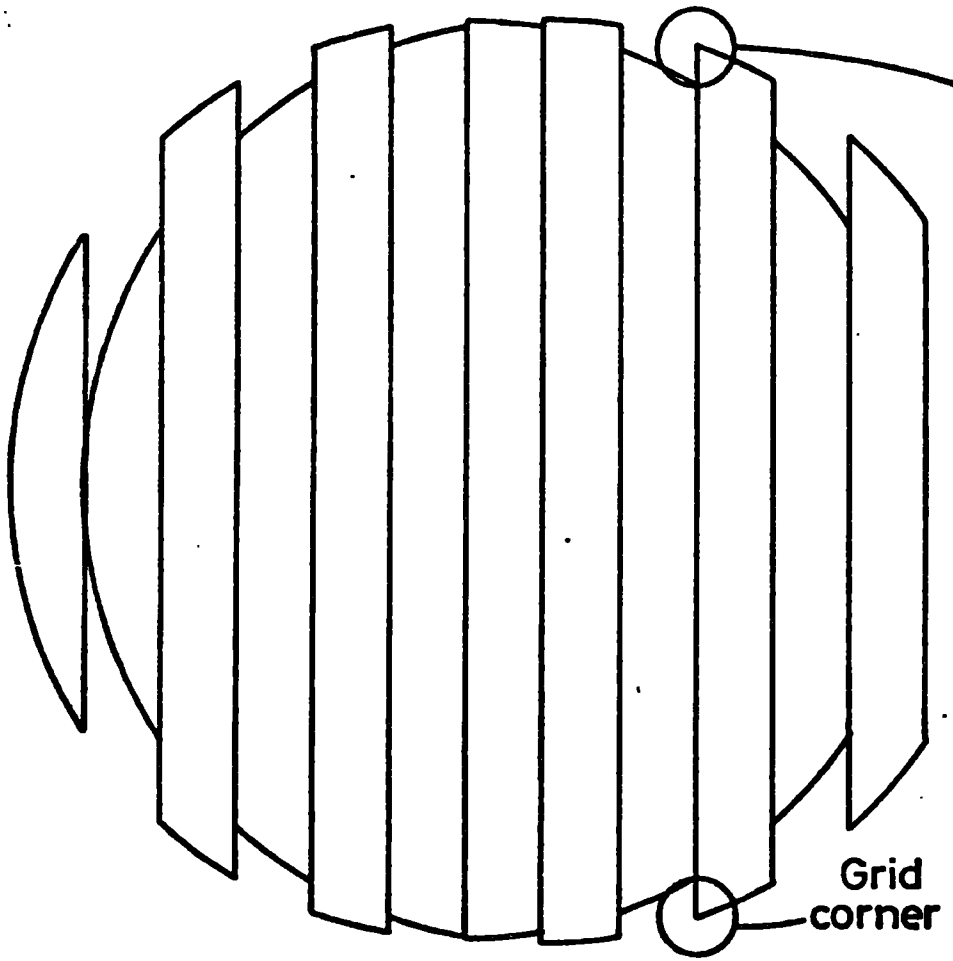
5.5.5 The Master Grid Spot Positions

The 4 sets of grid spot positions are aligned with each other by transforming 3 of the sets to co-incide with the 4th. This is done using the usual 6 parameter linear transformation as used for aligning the individual electronographs when making the photocathode correction (5.2.5), the same rejection technique for checking the validity of the positions found is also used.

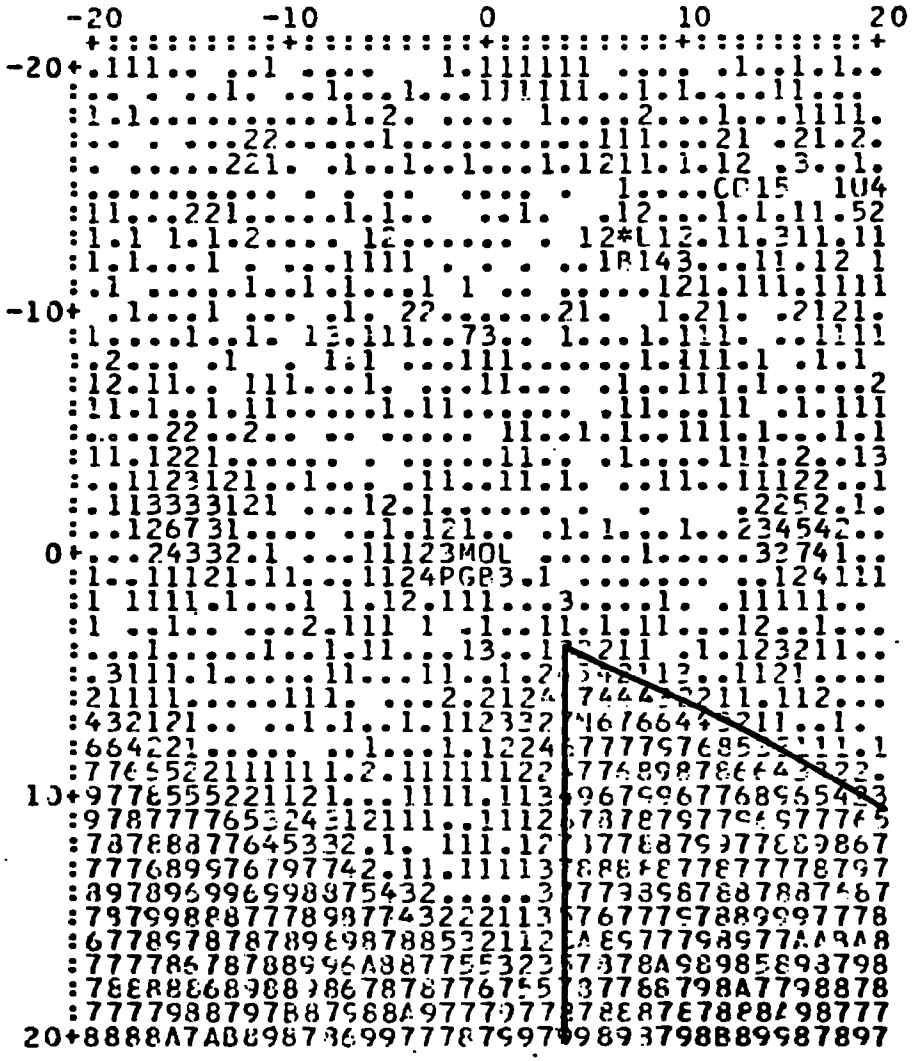
Having obtained 4 sets of aligned grid spot positions, the positions of the grid corners (5.5.4) are transformed in the same way, and all 4 sets of valid grid spot and grid corner positions are averaged to give mean grid spot and grid corner positions.

Finally, the set of mean positions is transformed again, by applying a rotation centred on the upper mean grid corner position, until the lower grid corner has the same X coordinate as the upper one. The resulting "master" grid spot positions are then such that any electronograph aligned with them will have its grid bar images parallel to the Y axis.

These positions are used for the alignment of all data reduced using this photocathode response and form part of the photocathode file. In order that data aligned



Polarimeter image



Digitised grid corner

Fig.5.18 Locating grid corners by eye

to these positions should also be aligned to the photocathode response, the electronographs used to produce the response are themselves aligned to these positions (5.5.7).

5.5.6 Determining the Left-Right Transformation

When aligning the data, we need to know (5.3.7) the relative positions of the left and right hand images in the frame of reference of the photocathode matrix. The transformation between these two positions can be found at this stage from the positions of the two images formed of each grid spot.

The form used for the transformation relates the difference in the positions of the two images $(\Delta X, \Delta Y)$ to the position of the left hand image (X_L, Y_L) using a 6 parameter linear transformation:

$$\Delta X = t_1 + t_2 X_L + t_3 Y_L \qquad 5.21$$

$$\Delta Y = t_4 + t_5 X_L + t_6 Y_L \qquad 5.22$$

This form allows the difference between the two image positions to vary systematically across the polarimeter field. Such a variation is encountered in practice because of the slightly changing deviation introduced by the Wollaston prism at differing positions in the field. Typically such differences do not amount to more than 2 or 3 pixels.

The transformation coefficients are found using a least-squares criterion, namely minimising:

$$f = \Sigma\left[X_L + t_1 + t_2X_L + t_3Y_L - X_R\right]^2 + \Sigma\left[Y_L + t_4 + t_5X_L + t_6Y_L - Y_R\right]^2 \qquad 5.23$$

where L and R refer to the left and right images of each grid spot and the sums are formed over the 10 pairs of master grid spot positions.

5.5.7 Initial Processing of the Photocathode Calibration Data

The master grid spot positions and the left-right transformation determined above are combined to form a "dummy" photocathode file, which contains, at this stage, no photocathode response. Using this dummy file, the programmes used for the normal data reduction can be run to extract the numerical data from the calibration electronographs. The sequence of data reduction proceeds initially as normal, performing clear plate subtraction, linearity corrections and alignment with the master grid spot positions as described in the earlier parts of this chapter. At this point, however, some changes to the normal procedure are made:

i) No photocathode response is applied.

ii) No star positions are found on any image, hence the alignment rests solely on the alignment of the grid spots on each electronograph with the master grid spot positions.

iii) The normal left-right transformation (5.5.6) is modified so that the displacements between left and right hand image positions ($\Delta X, \Delta Y$) take on integer values. In this way it is possible to later re-construct the original set of 4 polarimeter images with the left and right images in their correct relative position within the matrix (rather than having 8 mutually aligned images as is normally required).

At the end of this initial processing stage, the data is in the form of 4 mutually aligned polarimeter images and we can proceed to the next stage of comparing these images to produce the photocathode calibration.

5.5.8 Determining the Grid Dimensions

In order to be able to process the images further, we must know the dimensions of each grid aperture within the matrix. The dimensions in the Y direction are found by inspecting line-printer pictures of the data, while the dimensions and positions in the X direction are found by examining traces made across the electronographs perpendicular to the grid bars. Figure 5.19 shows such a trace, and the areas of image overlap and gaps between images can be seen. The dimensions chosen for each aperture

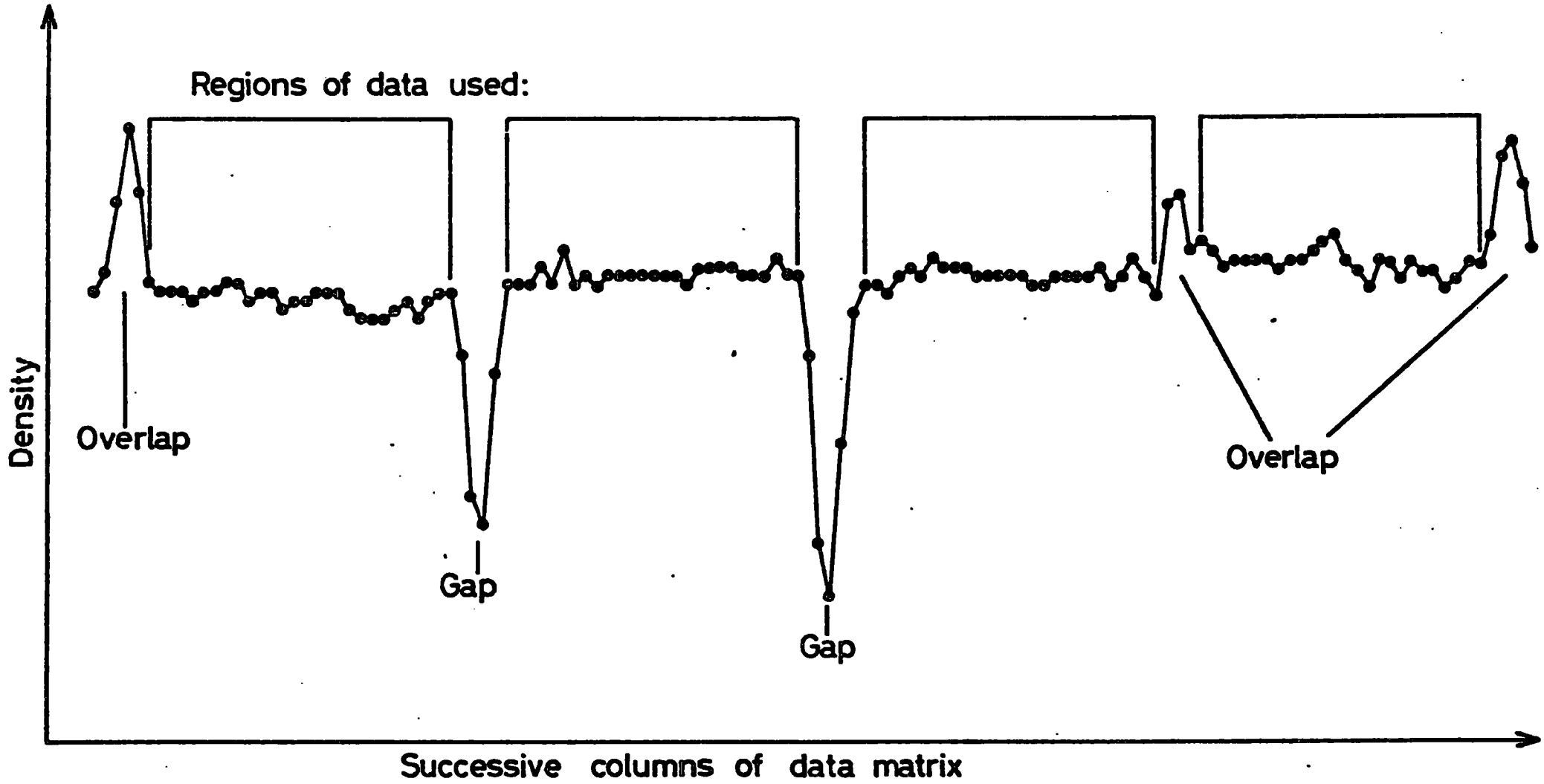


Fig.5-19 Determining the X dimensions of the grids

include only those areas which are free from such effects. Regions outside those chosen are flagged in the photocathode response as invalid, so that all data recorded in these regions is rejected when the photocathode response is corrected for (5.2.6, 5.2.7).

The regions regarded as valid at this stage are noted in the photocathode file, and this file forms a reference for the next stages of processing. From now on, the left and right hand images (6 apertures in each) are treated separately and identically, and the procedure will be described only once. No normalisation of the response between the left and right hand positions on the photocathode is necessary at this stage, as this is accounted for by the scale factors to be described in the next chapter.

5.5.9 Rejecting Bad Data by Smoothing

Before any comparison is made between the images recorded on each electronograph, each image is cleaned of defects to some extent by rejecting those pixels which deviate from the local smoothness within each image. The procedure is as follows:

A smoothed version of each image is formed by replacing each pixel with the average of the 11 x 11 pixel area centred on it. A comparison is then made between the smoothed and unsmoothed versions (S_j and U_j) of each pixel to estimate the RMS noise present on the

image (assumed constant, since the density recorded is very nearly uniform):

$$\sigma = \sqrt{\frac{1}{N} \sum_{j} (S_j - U_j)^2} \qquad 5.24$$

In estimating the noise, those pixels which have $|S_j - U_j|$ of more than 50 PDS units are not used. N is the total number of pixels which are used.

Having found an estimate of the noise, those pixels whose deviation from the smoothed value is more than 2σ are rejected. The noise is then re-evaluated to give a more accurate estimate.

5.5.10 Scaling the Data for Comparison

Since each of the 4 electronographs used will have been exposed to a slightly different density, they must be normalised to each other before a comparison can be made. To do this accurately, identical regions on each electronograph must be used in the normalisation. This is achieved by selecting those pixels which have not been rejected on any of the 4 images and calculating the mean density on each image for these pixels. The individual electronographs are then scaled to make their mean densities identical, and the estimates of the RMS noise are scaled likewise. We then have, at each point, up to 4 comparable measurements of the photocathode sensitivity d_k (k=1,4) and associated estimates of the errors σ_k.

5.5.11 Comparing the 4 Measurements

At each point within the images, a "weighted mean" estimate of the photocathode sensitivity is made:

$$\langle d \rangle = \frac{\sum_k \frac{d_k}{\sigma_k^2}}{\sum_k \frac{1}{\sigma_k^2}} \qquad 5.25$$

Points which have been rejected are omitted when forming the sums. To test the validity of the (up to 4) measurements, the deviation from this mean is compared with that expected on statistical grounds.

The expected dispersion of the quantity $d_k - \langle d \rangle$ can be shown to be:

$$\tau_k = \sqrt{\left[\sigma_k^2 + \frac{1}{\sum_j \frac{1}{\sigma_j^2} - \frac{1}{\sigma_k^2}}\right]} \qquad 5.26$$

where $\sum_j$ is a sum over all the measurements used.

Because only 4 measurements are used, the calculated mean can be highly affected by only one bad measurement. To prevent this causing the unnecessary rejection of good data, a method similar to that used when aligning grid spots is used. The seriousness of the deviation of each measurement from the mean is found from:

$$\chi_k^2 = \frac{(d_k - \langle d \rangle)^2}{\tau_k^2} \qquad 5.27$$

and the maximum value of χ^2_K is found (i.e. that measurement with the most serious deviation). If this measurement has χ^2_{MAX}, it is rejected if:

$$\chi^2_{MAX} > 4 \qquad 5.28$$

and the process is repeated, omitting this point. A calculated mean value is only accepted if $\chi^2_{MAX} < 4$. If this cannot be achieved until only one measurement is left - indicating complete disagreement between all 4 electronographs - no measurement is used at that point and the photocathode response is flagged as invalid.

Finally, the weighted mean is used to represent the photocathode response, provided it lies within $\pm$30% of the mean value to which the individual electronographs have been normalised. In this way "hot spots" and "dead spots" are rejected.

5.5.12 The Photocathode File

The photocathode file contains those quantities, relevant to the reduction of all data, which have been calculated at the stage of photocathode calibration. They are:

i) The (X,Y) positions of the 2 master grid corners. (5.5.5)

ii) The (X,Y) positions of the 20 master grid spots. (5.5.5)

iii) 6 left-right transformation coefficients. (5.5.6)

iv) The X positions of each edge of the 6 left hand

grid aperture images.

v) The y positions of top and bottom of the 6 left hand grid aperture images.

vi) A 384 x 384 matrix of integers representing the photocathode response and containing flags to indicate invalid regions.

A typical photocathode response produced by this method is shown in figure 5.20, where the symbols (in alphabetical order of increasing sensitivity) represent 2% changes in sensitivity with respect to the mean. Typical variations across the polarimeter field amount to $\pm$2 to $\pm$10%, usually with maximum sensitivity near the centre of the field.

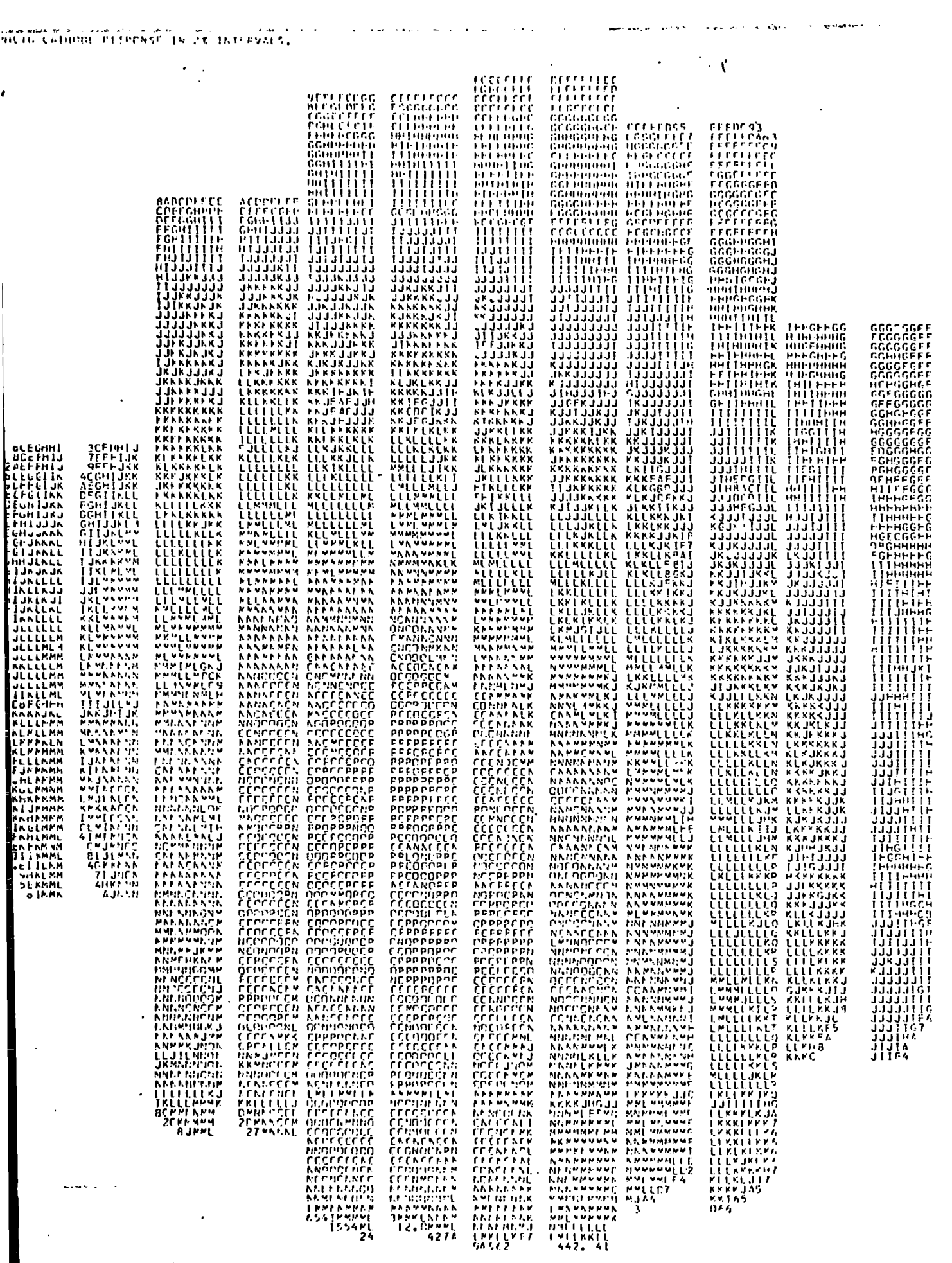


Fig. 5.20 Typical photocathode calibration

CHAPTER 6

SCALING THE DATA AND CHECKING LINEARITY

6.1 Introduction

So far we have dealt with the extraction of the numerical data from the electronographs in a form suitable for further processing. Some such technique is required whenever numerical information is needed from electronography, and the complexity of the technique used here reflects the fact that the later stages of data-reduction provide many checks on the accuracy with which this has been accomplished.

It is to these stages which we now turn our attention, and particularly to checking the linearity of response in this chapter.

6.2 The E and F Scale Factors

6.2.1 The Role of the Scale Factors

As was shown in the simple theory of the reduction technique described earlier (4.1), there are 4 scale factors, E_1, E_2, E_3, F which must be determined and applied as corrections to the measured densities d_1 to d_8 before the polarisation can be calculated. These scale factors are, in turn, derived from the measured densities.

Earlier reduction techniques (Axon 1977, Pallister 1976) adopted the simple approach of evaluating the scale

factors at each point in the image from the expressions 4.14 to 4.19. There are a number of disadvantages to this method which have led to its replacement:

i) The ratios calculated are extremely sensitive to errors in the background subtraction; gross errors resulting from only small errors in sky signal subtraction.

ii) Because the densities d_1 to d_8 are corrected with scale factors which depend in turn on d_1 to d_8, the corrected values are not mutually independent and the calculation of errors is complicated.

iii) If, for any reason (e.g. corrupt data), one of the densities d_1 to d_8 cannot be used at any point, it becomes impossible to calculate one of the scale factors and hence some of the remaining 7 densities cannot be corrected and must also be rejected.

To overcome these problems, the calculation of the scale factors at each point has been replaced with the calculation of a single set of scale factors applicable to the entire image. This is only possible after an accurate photocathode calibration has been carried out. Problems (ii) and (iii) are then resolved, since the scale factors are determined from the entire image and are largely independent of any individual pixel, and problem (i) is resolved by the new method used to determine the scale factors (6.2.3).

6.2.2 The Need for Accurate Scale Factors

To a good approximation, the calculation of the

scale factors is barely necessary since they have only a small effect on the polarisation calculated. (For example, a 10% error in E_1 would mean that 10% polarisation would be measured as 10.25%, while a 10% error in F would give 10.0% - the correct result).

This is not true, however, if some of the 8 intensities d_1 to d_8 are invalid and omitted from the calculation. In this case, the cancellation in the equations which makes the scale factors unimportant no longer takes place and accurate scale factors are needed.

A further factor is the need to make estimates of the errors on the final results. This is done by checking the consistency of the 8 densities, and if they are wrongly corrected at this stage, the errors calculated will be considerably larger than the true errors. As a result, we aim to determine the scale factors to within 0.5%.

6.2.3 Determination of the Scale Factors

Consider the equations 4.1 to 4.13, presented earlier, but consider now, also, the possible residual background signals s_1 to s_8 due to imperfect sky subtraction:

$$d_1 = i_1 + s_1 \qquad 6.1$$

$$d_2 = Fi_2 + s_2 \qquad 6.2$$

$$d_3 = E_1 i_3 + s_3 \qquad 6.3$$

$$d_4 = E_1 F i_4 + s_4 \qquad 6.4$$

$$d_5 = E_2 i_5 + s_5 \qquad 6.5$$

$$d_6 = E_2 F i_6 + s_6 \qquad 6.6$$

$$d_7 = E_3 i_7 + s_7 \qquad 6.7$$

$$d_8 = E_3 F i_8 + s_8 \qquad 6.8$$

also:

$$i_1 = i_6 \qquad 6.9$$

$$i_2 = i_5 \qquad 6.10$$

$$i_3 = i_8 \qquad 6.11$$

$$i_4 = i_7 \qquad 6.12$$

$$i_1 + i_2 = i_3 + i_4 = i_5 + i_6 = i_7 + i_8 \qquad 6.13$$

If we now form straight-line plots as follows, by plotting all the pixels of one image against the corresponding ones of another image, we can find the intercepts and gradients from equations 6.1 to 6.13:

Plot (Y)	Against (X)	Gradient	Intercept on Y axis
d_6	d_1	$G_1 = E_2 F$	$S_6 - s_1$
d_5	d_2	$G_2 = E_2/F$	$s_5 - s_2$
d_8	d_3	$G_3 = E_3 F/E_1$	$s_8 - s_3$
d_7	d_4	$G_4 = E_3/(FE_1)$	$s_7 - s_4$

(Table 1)

Hence we can find two estimates of F, independent of s_1 to s_8:

$$F_1 = \sqrt{G_1/G_2} \qquad 6.14$$

and

$$F_2 = \sqrt{G_3/G_4} \qquad 6.15$$

Using the mean of these two estimates, we can now form the following plots:

Plot (Y)	Against (X)	Gradient	Intercept on Y axis
$d_3 + d_4/F$	$d_1 + d_2/F$	E_1	$(S_3+S_4/F) - (s_1+s_2/F)$
$d_5 + d_6/F$	"	E_2	$(s_5+s_6/F) - (s_1+s_2/F)$
$d_7 + d_8/F$	"	E_3	$(s_7+s_8/F) - (s_1+s_2/F)$

(Table 2)

Hence the E factors can also be found independently of s_1 to s_8. Furthermore, we have a check on the accuracy of the sky subtraction since all the plots should have zero intercepts if the sky subtraction were perfect.

6.2.4 Binning the Data for the Plots

Because the above plots contain $\sim$30,000 points, it is very time consuming to use a normal least-squares fit to determine the gradients and intercepts, consequently a binning technique is employed.

The points are divided into 250 bins corresponding to equal increments of the abscissa. In each bin the mean value of the ordinate and abscissa is found, resulting in 250 X-Y pairs to which the fit is made. Because of the presence of bad data, however, several stages of data rejection are necessary to obtain the required accuracy. The first stage of rejection removes points before they are included in the bins, as follows.

When the mean ordinates are calculated above, the standard deviation of the ordinates within each bin is also calculated. The binning process is then repeated, but only points with ordinates within 2 standard deviations of the mean ordinate are accepted into each bin. The mean and standard deviation are then re-calculated and the binning process repeated again to yield a final mean ordinate for each bin. An estimate of the probable error, ε, on the mean is also made from the standard deviation and the number of points within the bin. Finally, bins containing less than 5 points are not used.

While the above procedure ensures that the binned quantities are not corrupted by spurious data points, it is still possible for entire bins to be in error. This often arises at high densities where the majority of points within a bin come from within star images, where the image alignment is not sufficiently accurate to compare individual pixels in different images. Because of this, a second stage of data rejection, involving whole bins, is used when the fits are made to the binned X-Y values.

6.3 The Linearity Errors

6.3.1 Introduction

When fits to the above binned quantities are made, it is usually found that the plots are not perfectly straight lines, and consequently it is not possible to

characterise the entire plot by a single scale factor applicable at all densities. The reason for this non-linearity is apparently the differing non-linearities in the different electronographs.

To overcome this problem, the scale factor E has been split into two parts: $\bar{E}$, which represents the best straight-line fit to the data, and e, which represents the deviations from the straight line and is an additive function of the density of the electronograph (Figure 6.1).

In order to obtain good correction of the data with the scale factors for the reasons outlined in section 6.2.2, it is necessary to determine, and correct for, both of these components.

6.3.2 Fitting the Linearity Deviations

The linearity deviations are found using the 3 plots in table 2 (6.2.3), with F set to unity at this stage with negligible loss of accuracy. The first step is to determine the factor E approximately by making a straight line fit to the data. A weighted least-squares criterion is used, chosing the parameters a and E to minimise:

$$f = \sum_i \frac{\left[Y_i - (a + EX_i)\right]^2}{(\varepsilon_i^2 + k^2)} \qquad 6.16$$

where the sum is over the binned values (X_i, Y_i) calculated above (6.2.4), and the weights are derived from the expected errors ε_i on the values Y_i. The constant

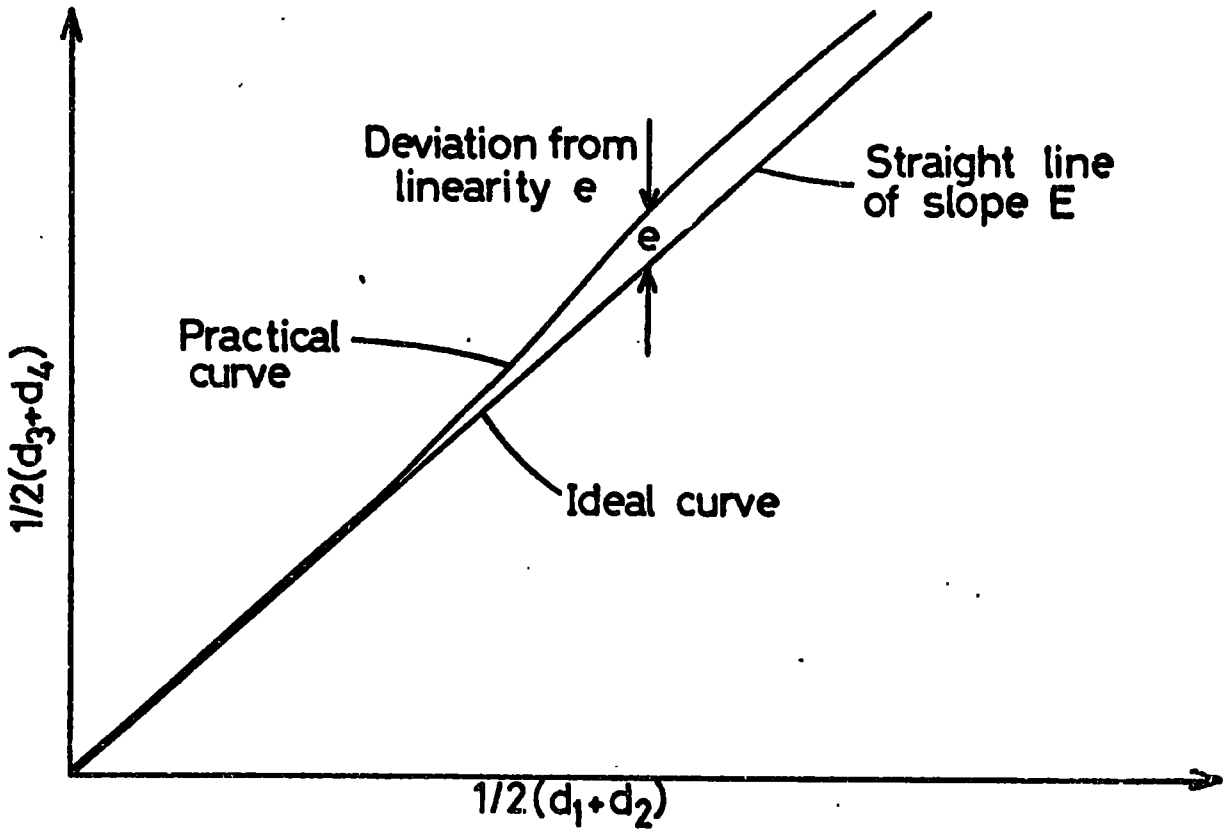


Fig.6-1 Two components of the 'E' factor

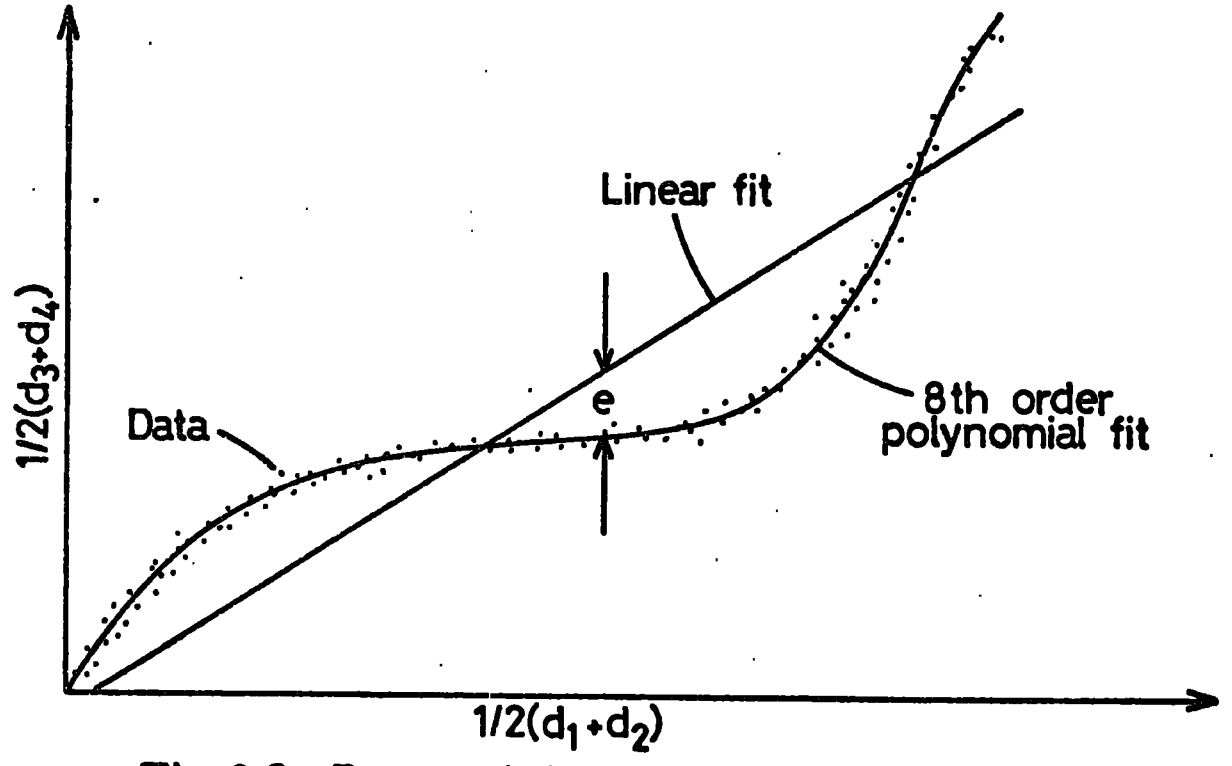


Fig.6-2 Determining the linearity deviations

k is included to prevent excessive weight being given to any bin, and is set at the approximate expected minimum error on a bin of 0.8 PDS units.

Having obtained one fit by the above method, those bins which have:

$$\frac{|Y_i - (a + EX_i)|}{\sqrt{(\varepsilon_i^2 + k^2)}} > 3 \qquad 6.17$$

are rejected, and another fit obtained. This rejection is made deliberately weak, since some points may deviate from the fit due to non-linearity rather than corrupt data.

Having obtained an approximate linear fit to the plot, the linearity deviations are found by making an 8th order polynomial fit. A modified least-squares technique is employed, by chosing a polynomial $P_8(X)$ which minimises:

$$f = \sum_i \frac{[Y_i - P_8(X_i)]^2}{(\varepsilon_i^2 + k^2)} + g \sum_{j=1}^{j=110} [P_8(hj) - P_8(h[j-1]) - hE]^2 \qquad 6.18$$

The first term is the normal weighted least-squares criterion. The second term is an empirically derived "smoothness" constraint to minimise rapid oscillation of the fitted polynomial, and to ensure that it represents a good extrapolation at the ends of the data. The constant h is chosen so that 111 equally spaced samples of the polynomial are taken and the mean gradient between samples

is compared to that of the straight line fit. In this way the fit is constrained to be as close to the straight line fit as possible, until deviated by the data points. The constant g determines the strength of the constraint.

Rejection of bad bins is accomplished as above, rejecting points which satisfy equation 6.17, and re-evaluating the fit. The linearity deviations are finally calculated as the difference between the 8th order polynomial fit and the straight line fit, and are expressed as 8th order polynomial functions of $(d_1 + d_2)$ (Figure 6.2). Figure 6.3 shows typical linearity deviations, where the straight line fit has been subtracted to show the data points also.

6.3.3 Finding the Linearity Corrections

If we are to make corrections to the data on the basis of the linearity deviations, we must assume that the deviations on any one electronograph are characterised by their density, and that the densities in the two images on each electronograph are sufficiently similar for them to have the same linearity errors. We can then find the linearity deviation for a given density by considering the average of the two images. Figure 6.4 shows that this is a good assumption since the deviations are slowly varying functions of density. The next step, therefore is to process the 3 linearity deviations e_1, e_2, e_3 which are functions of $(d_1 + d_2)$ into 4 corrections to be applied to the 4 electronographs. Obviously this cannot be done uniquely and some form of average linearity must be adopted.

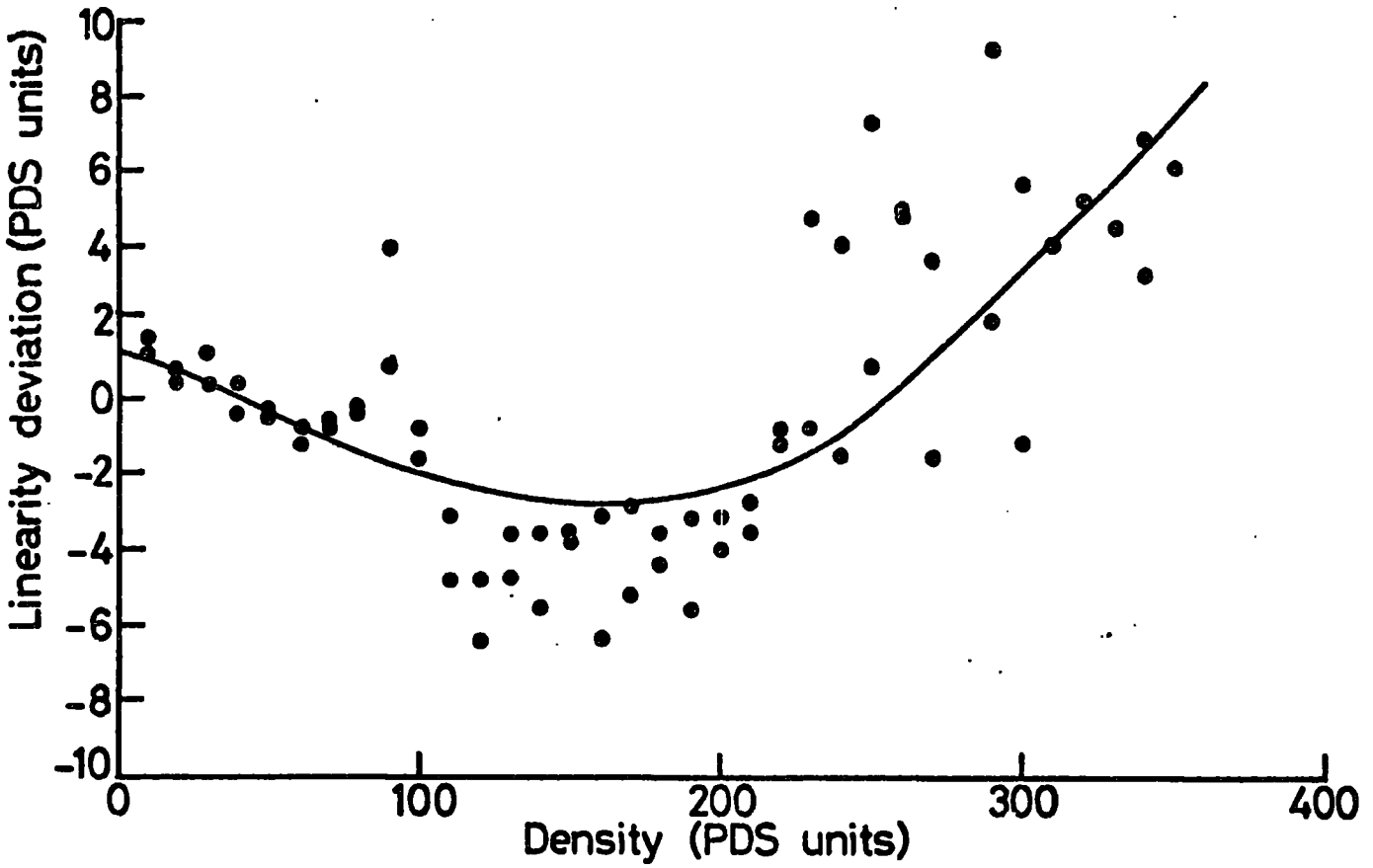


Fig.6-3 Typical linearity deviations

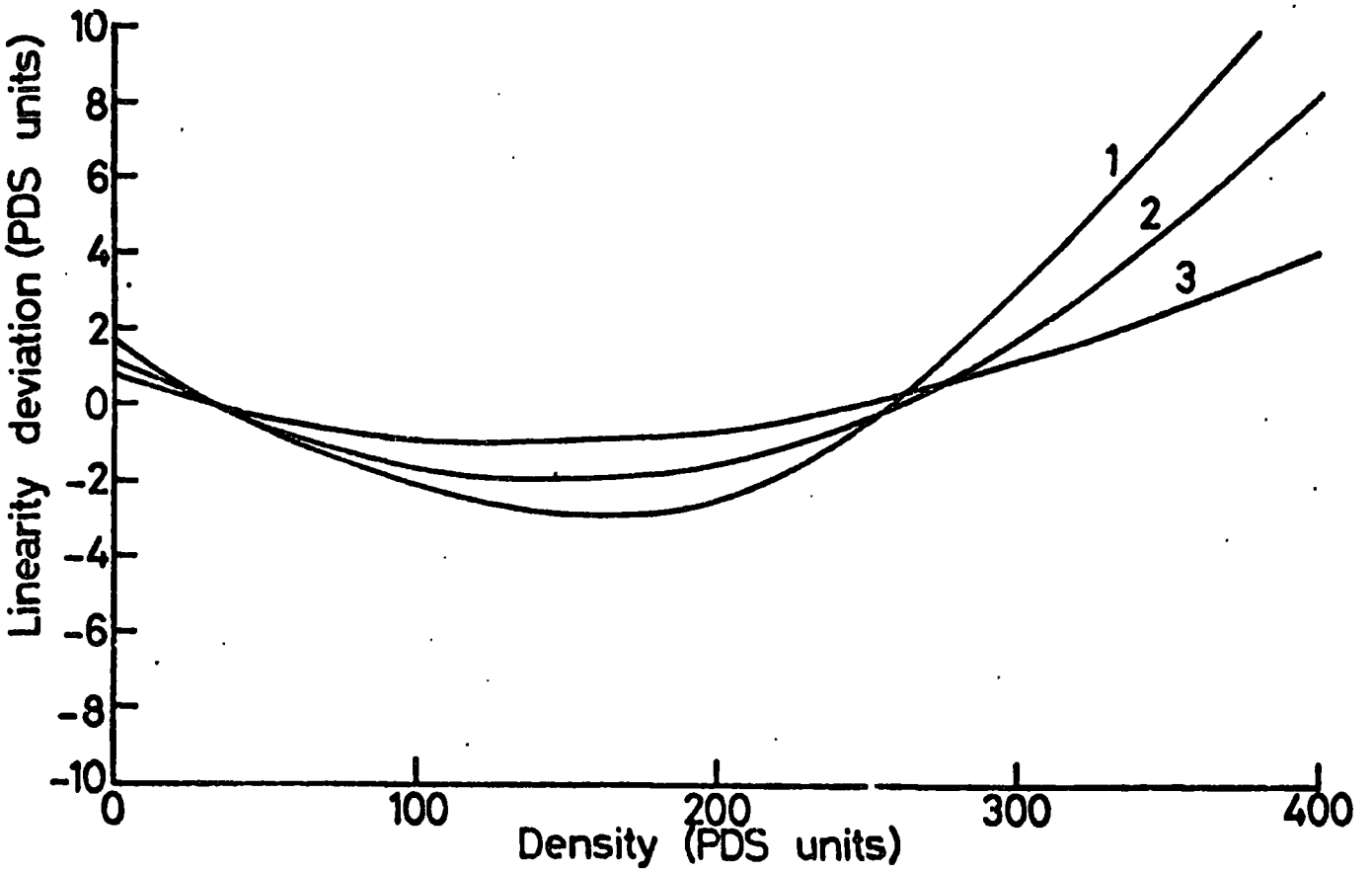


Fig.6-4 Linearity deviations on E_1, E_2, E_3 for a typical set of data

Suppose that at corresponding densities on the 4 plates (i.e. densities related by the straight line fits), the linearity errors are Δ_i (i = 1,4). Then figure 6.5 shows that the deviations are given by:

$$e_1 = -E_1 \Delta_1 + \Delta_2 \qquad 6.19$$

$$e_2 = -E_2 \Delta_1 + \Delta_3 \qquad 6.20$$

$$e_3 = -E_3 \Delta_1 + \Delta_4 \qquad 6.21$$

To impose the condition of "average" linearity we also make:

$$\Delta_1 + \Delta_2 + \Delta_3 + \Delta_4 = 0 \qquad 6.22$$

(the sum of the deviations from the mean is zero), and solving these last 4 equations for the Δ_i we have:

$$\Delta_1 = \frac{-(e_1 + e_2 + e_3)}{1 + E_1 + E_2 + E_3} \qquad 6.23$$

$$\Delta_2 = e_1 + E_1 \Delta_1 \qquad 6.24$$

$$\Delta_3 = e_2 + E_2 \Delta_1 \qquad 6.25$$

$$\Delta_4 = e_3 + E_3 \Delta_1 \qquad 6.26$$

Hence, knowing e_1, e_2, e_3 as functions of the density on the first electronograph, and also having straight line relationships between the densities on each electronograph, we can find the corrections $(-\Delta_1)$ to be made to

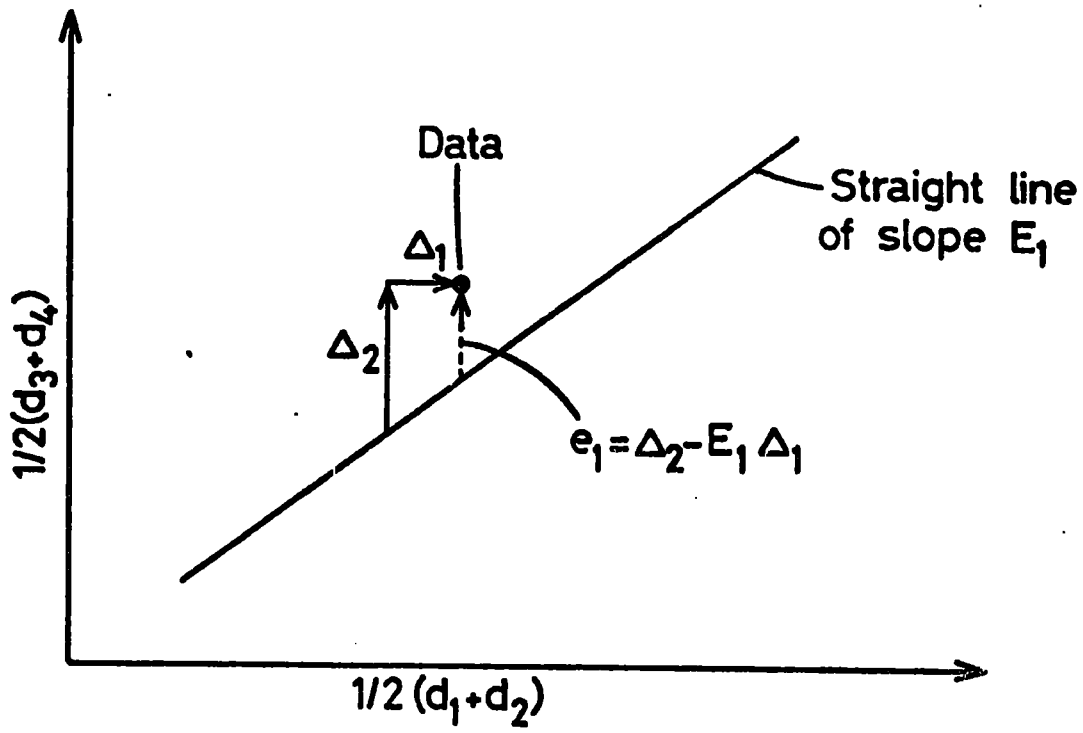


Fig. 6.5 Determining the linearity corrections

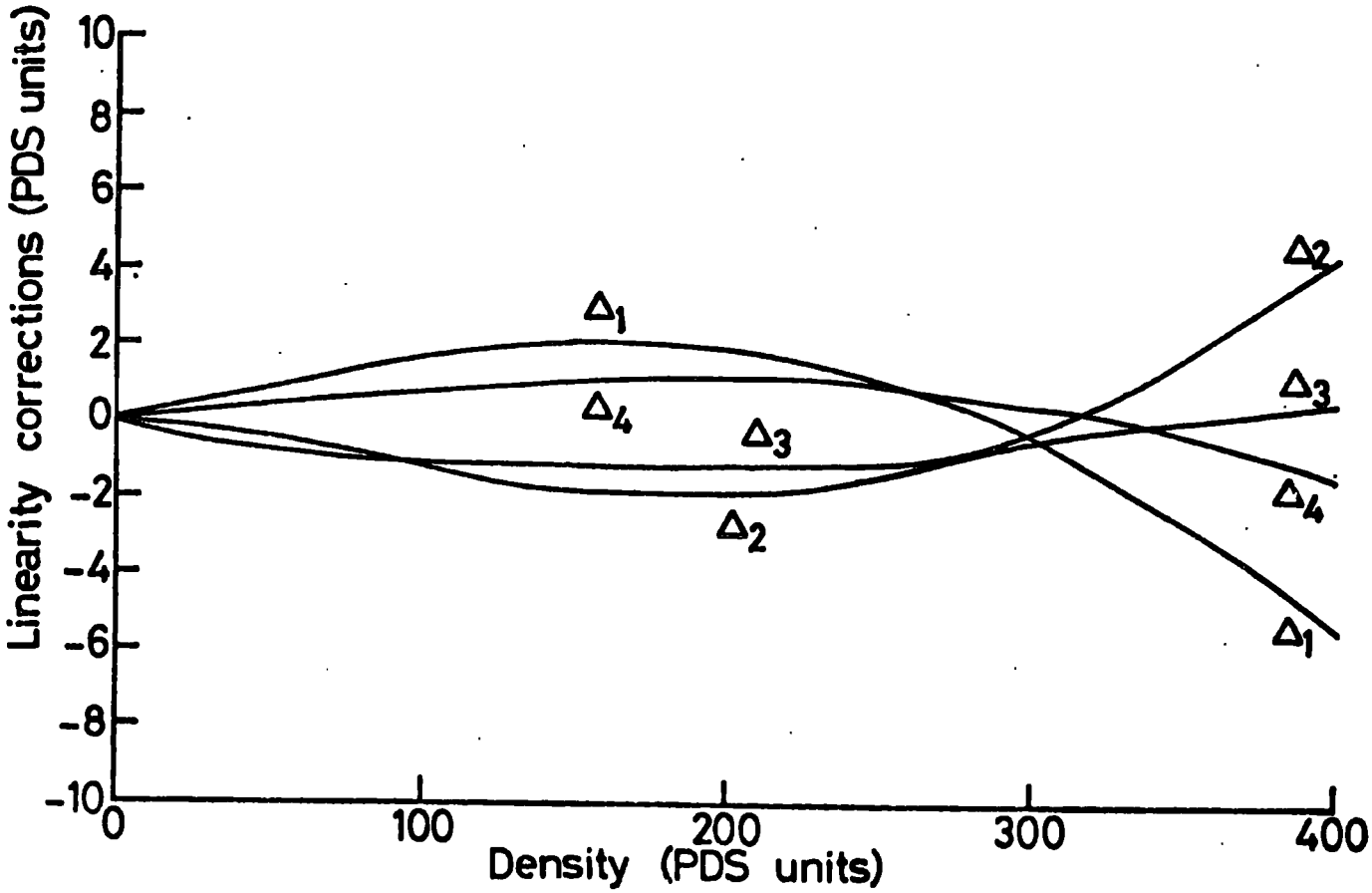


Fig. 6.6 Corrections derived from figure 6.4

the 4 electronographs as functions of the density on the electronographs to be corrected.

Finally, each of the corrections derived is altered by the addition of a constant such that no correction is made when the density is zero. This is necessary in order that the zero level of density, which is determined by the sky signal subtraction, is not altered before checks can be made on the accuracy of the sky subtraction. Figure 6.6 shows the linearity corrections derived from the deviations in figure 6.4.

6.4 Making the Corrections

6.4.1 The Linearity and Scale Factor Corrections

Having determined the linearity corrections for the 4 electronographs, the data is corrected with them, as are the binned quantities used for the scale factor plots. Since the linearity deviations represent deviations from a straight line, all these plots should now be straight lines and the scale factors can be determined using the sequence described in section 6.2.3. In determining the straight line fits to the plots, the same technique is used as in 6.3.2 (Equation 6.16), except that the rejection procedure is made more strict, points with

$$\left| \frac{Y_i - (a + GX_i)}{(\varepsilon_i^2 + k^2)} \right| > 2 \qquad 6.27$$

(where G is the gradient of the line) being rejected after the first fit has been found.

Finally, having determined the scale factors by this technique, the data is corrected with them to convert d_1 to d_8 into i_1 to i_8, ready for calculation of the polarisation parameters.

6.4.2 Accuracy

There is no direct check on the accuracy of the E factors. However, some indication of accuracy can be obtained by comparing the two estimates of the F factor obtained (6.2.3). Typically these agree within ~0.5% and differ from unity by 1 or 2%.

Another check is to compare the values of F found for different sets of electronographs using the same photocathode correction, since F should depend only on the photocathode correction and the polarimeter optics. Agreement is usually found to within 1 or 2%, but it is highly dependent on the quality of the data, particularly the range of density present on the electronographs and the number of bright stars in the field.

In conclusion, it seems that the accuracy aim of 0.5% has only been achieved under conditions of near-ideal data, the main problem still being the presence of corrupt data points. However, the accuracy is not significantly worse than the aim in most cases, and further improvements to the technique have not been thought worthwhile.

6.4.3 Sky Subtraction Checks

As pointed out in section 6.2.3, the intercepts in the scale factor plots give an indication of the sky subtraction accuracy. In practice, these intercepts are usually less than 1 PDS unit, and this seems to represent the intrinsic uncertainty in the determination of the intercepts ($\sim\pm$ 0.5 PDS units). Consequently, sky subtraction errors can usually be detected if they amount to substantially more than 1 PDS unit.

6.4.4 Storing the Linearity Errors

As part of the error calculations (to be described fully in chapter 7), the linearity errors derived here are used as an indication of the errors in density measurement as a result of non-repeatable linearity of the electronographs. Of course, this does not represent the total error from non-linearity, since all 4 electronographs could be similarly non-linear, and thus in mutual agreement. However, linearity corrections have been applied (5.1.6) for all the known causes of non-linearity, while non-repeatable effects cannot be corrected for and must necessarily give an unpredictable error in the result. It is this unpredictable component which we are interested in here.

An estimate of the RMS density error from this cause at a given density is:

$$\langle \Delta \rangle = \sqrt{\sum_{i=1}^{4} \Delta_i^2 / 3} \qquad 6.28$$

(assuming the scale factors are unity, which is sufficiently accurate here). Consequently, the probable error on the mean of 4 electronographs is:

$$L = \langle \Delta \rangle / \sqrt{4} \qquad 6.29$$

This quantity is stored as a function of density to be used later in the error calculations.

CHAPTER 7

ERRORS, DATA VALIDATION AND CALCULATION OF THE POLARISATION PARAMETERS

7.1 Introduction

In the earlier stages of the reduction technique, the presence of corrupt data has been a major obstacle to accurate results. In each case, a solution has been adopted which is specific to the problem in hand, but none of these solutions lends itself to the detection of all the corrupt data present while minimising the loss of valid data.

When calculating the polarisation parameters, it is essential that no corrupt data is present. The prime reason for this is that regions containing many pixels (typically 20 to 30) are averaged when the polarisation is calculated, and the probability of there being no corrupt data, on any of the 8 images, within such an area, is small. As a result, only a small fraction of the polarisation measurements will be without serious error if the usual density of corrupt data is allowed to remain.

Because of this difficulty, a technique has been developed to check the entire set of data for complete internal consistency, and to reject corrupt data while minimising the loss of valid data. At the same time, valuable information on the other sources of error in

the data is obtained. Before the technique can be described, however, the nature and magnitude of these other sources of error must be considered.

7.2 Sources of Error

7.2.1 Photon Noise

The fundamental limitation on accuracy lies in the amount of light received from the object being studied. If we consider the detector as a counter of photons, then if N photons are detected, there must be an intrinsic uncertainty in this number of $\sqrt{N}$, resulting from the Poisson distribution of the number of randomly arriving photons detected in a given time. Thus there is "photon noise" on the result arising from the quantum nature of light. We must now consider the magnitude of this effect in electronography.

In the electronographic camera, not all the incident photons produce electrons which give tracks in the emulsion. This loss of information results in an increase in photon noise. Since, however, the number of incident photons is not usually known, we approach the problem from the other end and consider the "electron noise", resulting from those photo-electrons which <u>are</u> detected. We can estimate their number from the density of the electronograph.

Suppose that N electrons per unit area give a density of 1 on the electronograph, then within a square pixel of side ℓ, with density D, there are $ND\ell^2$ electrons.

Thus the electron noise within this pixel is $\sqrt{ND\ell^2}$ electrons.

Since each electron produces a large number of developable grains in the emulsion, the "grain noise" - that uncertainty in photographic density resulting from the discrete nature of the developed grains - will always be much less than the density uncertainty resulting from the electron noise. Consequently we neglect grain noise and determine the effect which the electron noise has on density. The uncertainty in electron density is $\sqrt{ND\ell^2}/\ell^2$, hence the uncertainty in photographic density on the electronograph is:

$$\Delta D = \frac{\sqrt{ND\ell^2}}{N\ell^2} = \frac{1}{\ell}\sqrt{\frac{D}{N}} \qquad 7.1$$

Thus we see that the "photon noise" error, due to those photons which <u>are</u> detected, varies as the square root of the recorded density. To find the magnitude of this effect, we use N = 4 electrons $(\mu m)^{-2}$ for L4 emulsion (RGO Electronographic Camera, Operator's Manual) and set ℓ = 25μm for the pixel size used. Using the conversion factor of 200 PDS units for a density of 1 we get:

$$\Delta D = \sqrt{0.08D} \qquad 7.2$$

where D is the density in PDS units. Substitution shows that for a typical "sky" density (20 PDS units), $\Delta D \sim 1.3$ PDS units, while for the upper limit of D = 1000,

$\Delta D \sim 9$ PDS units.

7.2.2 Quantisation Noise

Another fundamental source of error in the system arises from the fact that the densities measured on the PDS machine are rounded to integer values before being recorded. The resultant rounding error represents "quantisation noise" on the signal.

If the signal were a smoothly varying function of position, the quantisation would introduce steps (Figure 7.1), resulting in an error of up to $\pm$ 0.5 PDS units, highly correlated between adjacent pixels. If, however, there is noise present on the signal (e.g. photon noise) which is significantly larger than the quantisation interval, then the signal is random with respect to the quantisation levels and the rounding errors are uncorrelated and uniformly distributed between $\pm$ 0.5 PDS units (Figure 7.2). As a result, if several adjacent measurements are averaged, the quantisation noise is reduced resulting in a measurement somewhat more accurate than the quantisation interval. (For a fuller discussion of how the introduction of random noise can reduce errors in quantised systems see, e.g. Billingsley, 1975).

For the values used in the polarimeter, the photon noise exceeds the quantisation noise at all densities typically used, so we may consider the quantisation noise as uncorrelated. Integration over the distribution in

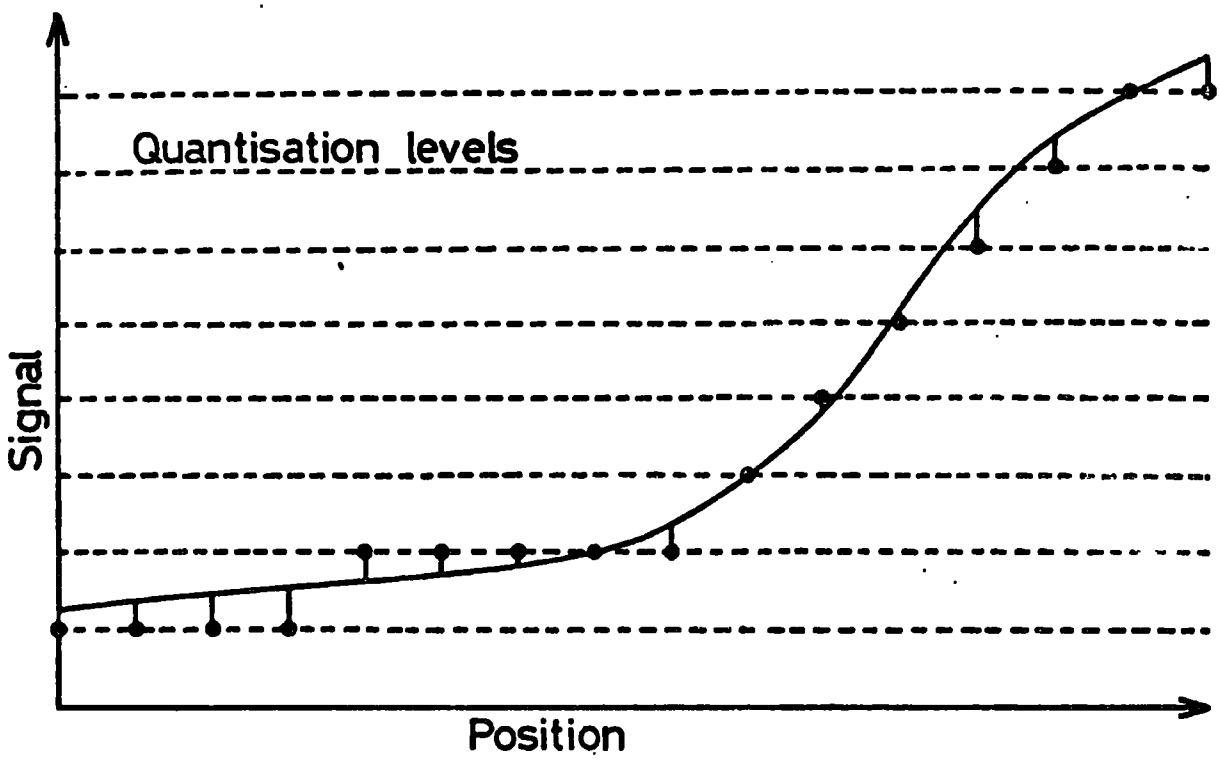


Fig.7.1 Quantising a smoothly varying signal

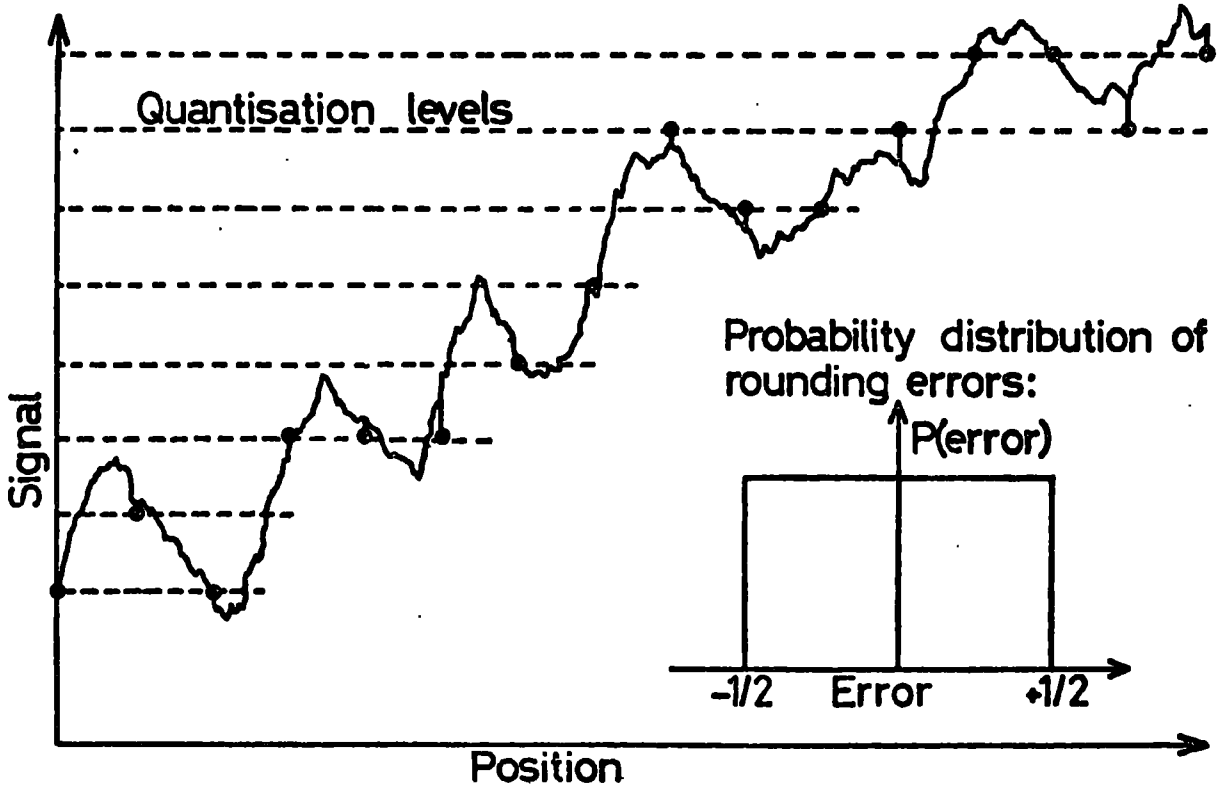


Fig.7.2 Quantising a noisy signal

figure 7.2 then gives the RMS quantisation noise as:

$$\Delta D = \frac{1}{\sqrt{12}} \quad \text{PDS units} \qquad 7.3$$

Combining this with the uncorrelated photon noise estimated previously (Equation 7.2) we have:

$$\Delta D = \sqrt{0.08D + 0.083} \qquad 7.4$$

as a fundamental limit to the accuracy of our measurements. The D term dominates this expression at all but the lowest densities, so that the quantisation introduces negligible error in practice.

7.2.3 Systematic Errors

The above errors will be referred to as "random", since they are to a good approximation uncorrelated between neighbouring pixels and the error on the result can be made small by averaging a sufficient number of pixels.

There are also errors, which will be referred to as "systematic which are correlated over large areas of the image, although they are largely un-correlated between the different electronographs. When such errors are present, increasing the number of pixels over which an average is taken cannot increase the accuracy of the result indefinitely - this can only be achieved by using more electronographs. Whenever the systematic errors

dominate, the practical accuracy of the system depends in a different way on the recorded densities than when random errors dominate. The combination of these two sources of error into an estimate of the total error will be discussed when the polarisation parameters are calculated.

The sources of systematic errors are manifold. In part, undetected dirt and scratches on the electronographs contribute, but there also appears to be a large contribution from the non-uniformity of the film used, both as regards emulsion quality and the transparency of the backing. There will be errors due to imperfect clear plate subtraction, photocathode calibration, and those due to non-linearity of response have already been discussed.

Although these errors come from a variety of sources, a large number of them have appreciable spatial extent on the electronographs and are thus highly correlated between the left and right image pairs. This correlation must be taken into account when the errors on the result are calculated.

7.3 The Theory of Data Validation

7.3.1 Introduction

The principle of data validation is to check the data for internal consistency, i.e. see that it has the form expected of it, and reject it if it does not. In

doing this, however, we must allow those errors which are acceptable (and will be averaged to give accurate results) to remain, while rejecting more serious corruption of data. The technique, therefore consists of two stages: i) Determining the form which the data should have, and ii) deciding whether the deviations from that form are acceptable or not.

7.3.2. The Functional Form of the Data

To accomplish the first stage, we need to find some function which describes ideal polarisation data. The data itself is a function of 3 variables: the X and Y position in the image and the orientation of the polarisation state into which the light is resolved. Thus our functional form should be a function of 3 variables (X,Y,θ) in order to describe as many aspects of the data as possible, and thus check for consistency as efficiently as possible.

The variation with orientation, θ, is determined by reference to Malus' Law describing the intensity of a beam of linearly polarised light when resolved into a linear polarisation state at an orientation θ. If the light is polarised at an orientation ϕ, the resultant intensity is:

$$i = \frac{I_o}{2} + I_p \cos^2(\theta - \phi) \qquad 7.5$$

where I_0 and I_p are the unpolarised and polarised intensities. From the definition of the Stokes parameters (I,Q,U) of the light, this can be written in the form:

$$i = \tfrac{1}{2}(I + Q \cos 2\theta + U \sin 2\theta) \qquad 7.6$$

The spatial variation of the data is, of course, not known. However, because of the effect of atmospheric "seeing", the recorded image is slightly blurred and adjacent pixels have non-independent signal values. Because of this correlation between adjacent pixels it is possible to use a Taylor series expansion to describe the local spatial variation with negligible error. The number of terms required in the series depends on the area of data to be represented, but for an area of 5 x 5 pixels, an expansion of the form:

$$I = I_1 + I_2X + I_3Y + I_4XY + I_5X^2 + I_6Y^2 \qquad 7.7$$

is sufficiently accurate for typical values of atmospheric seeing and for the image scales commonly used.

Expansions can also be used for Q and U, but fewer terms are needed for the same absolute accuracy since they are usually much smaller quantities than I. Hence, combining these spatial variations with the θ variation (Equation 7.6), we have a theoretical functional form for the data:

$$i_{TH}(X,Y,\theta) = I_1 + I_2X + I_3Y + I_4XY + I_5X^2 + I_6Y^2 + (Q_1 + Q_2X + Q_3Y)\cos 2\theta + (U_1 + U_2X + U_3Y)\sin 2\theta \qquad 7.8$$

The procedure is to determine the unknown constants in this expression from the local form of the data, and use the resulting function to predict an "interpolated" value at each point in the image. The actual data values can then be compared with this interpolated value to check their consistency with their neighbours in (X,Y,θ) space. The description of the data provided by this functional form is illustrated in figure 7.3.

7.3.3 Fitting the Functional Form to the Data

The functional form above must be fitted to a local region of data consisting of a 5 x 5 pixel region on each of 8 images, corresponding to 200 data points. Moreover, this must be repeated at every point in the image to determine the "theoretical" or interpolated value. Consequently, a prime consideration is the reduction of computing time to a minimum. The following analysis shows that so long as full advantage can be taken of the symmetry present in the data, some simplification is possible.

Using a least-squares fit, we want to find $i_{TH}(X,Y,\theta)$ which minimises:

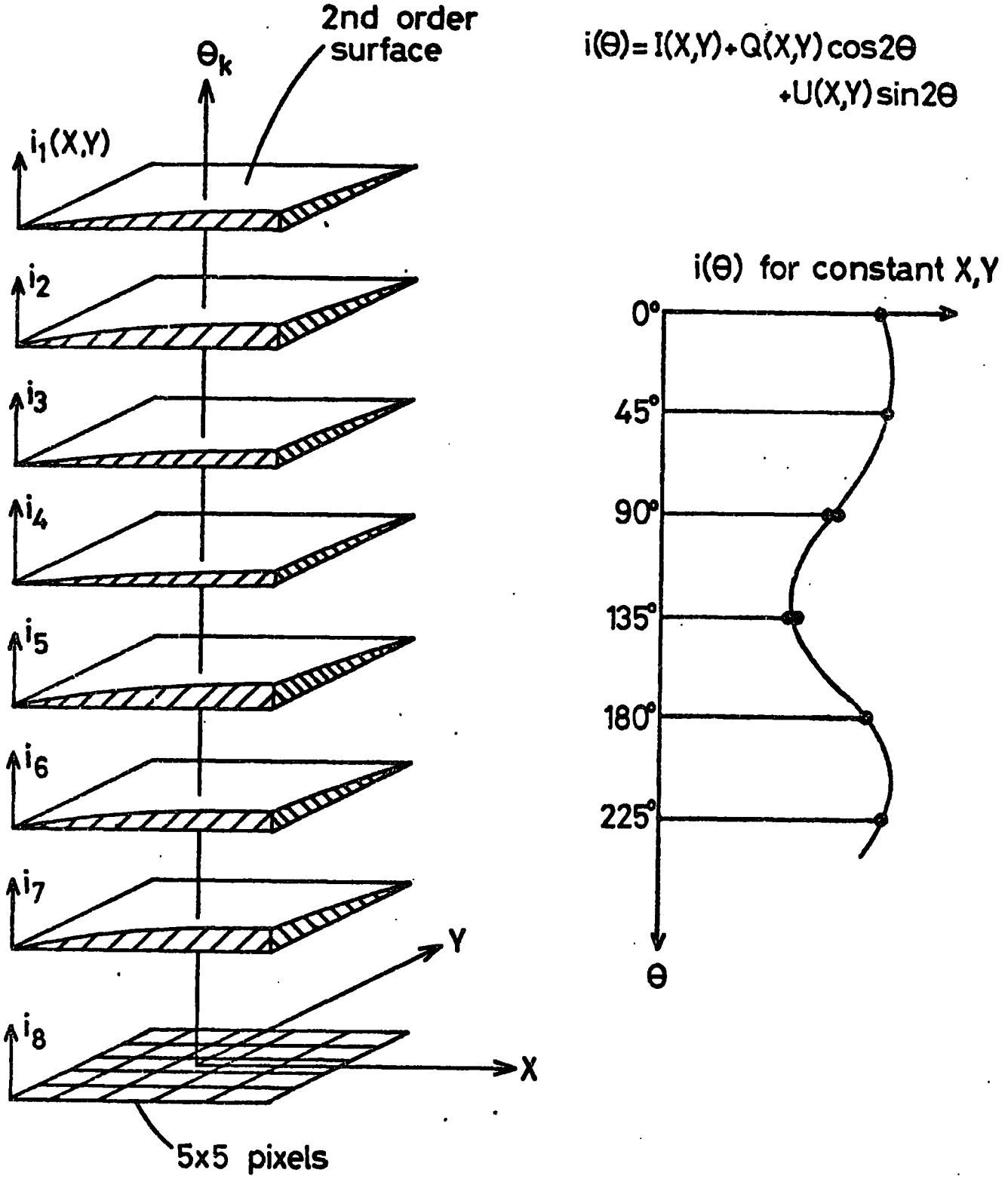


Fig.7.3 The form of the 3-D fit to the data

$$f = \sum_{j=1}^{5} \sum_{k=1}^{5} \sum_{\ell=1}^{8} \left\{ i_{TH}(X_j, Y_k, \theta_\ell) - i(X_j, Y_k, \theta_\ell) \right\}^2 \qquad 7.9$$

where the i's are the data points, and the sums are formed over the 200 data points shown in figure 7.3.

Substituting the expression for i_{TH}, differentiating in turn with respect to I_1, I_2, I_3, I_4, I_5, I_6, Q_1, Q_2, Q_3, U_1, U_2 and U_3 and equating these expressions to zero leads to a set of 12 simultaneous equations - known as the normal equations - which must be solved to give the I's, Q's and U's required for the least-squares fit. If we chose the origin of coordinates as in figure 7.4, and find our interpolated value at the central pixel, we need only evaluate i_{TH} at the origin, and hence we need only find I_1, Q_1 and U_1. The normal equations containing these variables are:

$$I_1 n + I_5 \Sigma X^2 + I_6 \Sigma Y^2 = \Sigma i \qquad 7.10$$

$$I_1 \Sigma X^2 + I_5 \Sigma X^4 + I_6 \Sigma X^2 Y^2 = \Sigma X^2 i \qquad 7.11$$

$$I_1 \Sigma Y^2 + I_5 \Sigma X^2 Y^2 + I_6 \Sigma Y^4 = \Sigma Y^2 i \qquad 7.12$$

$$Q_1 \Sigma \cos^2 2\theta = \Sigma \cos 2\theta . i \qquad 7.13$$

$$U_1 \Sigma \sin^2 2\theta = \Sigma \sin 2\theta . i \qquad 7.14$$

where n is the total number of data points (=200) and the sums are over all 3 variables (X,Y,θ). Use has been made of the fact that sums over terms containing

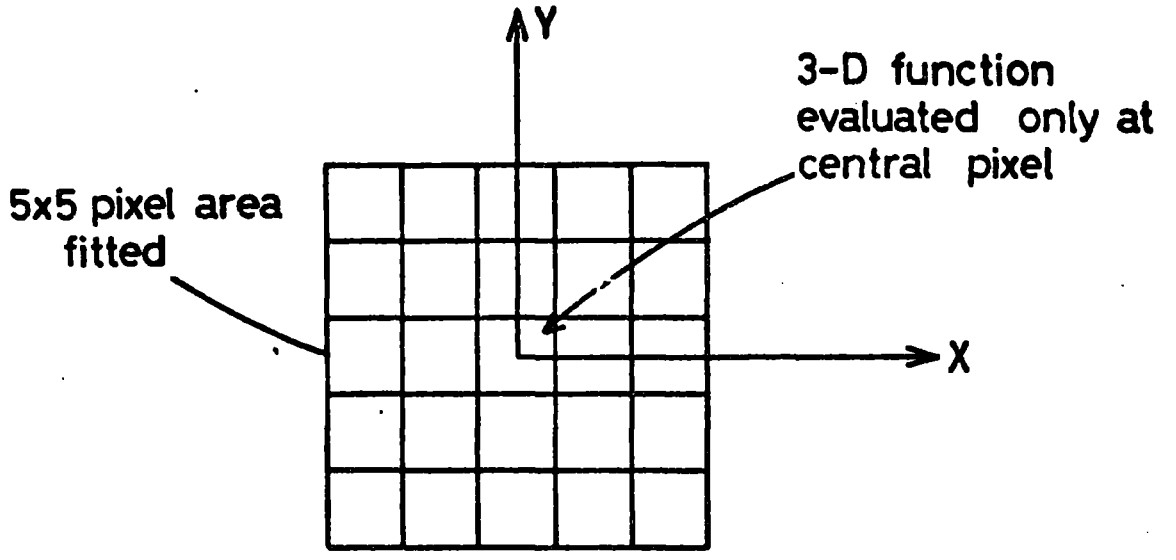


Fig.7.4 The coordinates used in finding the fit

k	θ_k	$\sin 2\theta_k$	$\cos 2\theta_k$	$\sin^2 2\theta_k$	$\cos^2 2\theta_k$
1	0°	0	1	0	1
2	90°	0	-1	0	1
3	45°	1	0	1	0
4	135°	-1	0	1	0
5	90°	0	-1	0	1
6	180°	0	1	0	1
7	135°	-1	0	1	0
8	225°	1	0	1	0

Fig.7.5 The trigonometric terms used in the fit

odd powers of X or Y are zero, because of the choice of origin, and also that:

$$\sum_k \sin 2\theta_k = \sum_k \cos 2\theta_k = 0 \qquad 7.15$$

for the values of θ_k used (Figure 7.5).

Solving equations 7.10 to 7.14 and substituting numerical values gives:

$$I_1 = 0.01929 \Sigma i - 0.003571 \, (\Sigma X^2 i + \Sigma Y^2 i) \qquad 7.16$$

$$Q_1 = 0.01 \Sigma \cos 2\theta . i \qquad 7.17$$

$$U_1 = 0.01 \Sigma \sin 2\theta . i \qquad 7.18$$

Hence the numerical work involved in finding the least-squares fit reduces to evaluating 5 summations over the 200 data points. The process can be speeded still further by making use of the overlap between adjacent fits (Figure 7.6) to eliminate much of the summation work.

7.3.4 Symmetry - Replacing Bad Points

Because the simplifications above are essential if the fitting process is to be sufficiently rapid to be useful, we must prevent the presence of invalid data from destroying the symmetry which enables these simplifications to be made. This problem also arises at the edges of the data, where the summation cannot extend over 200 data points because the data ends within

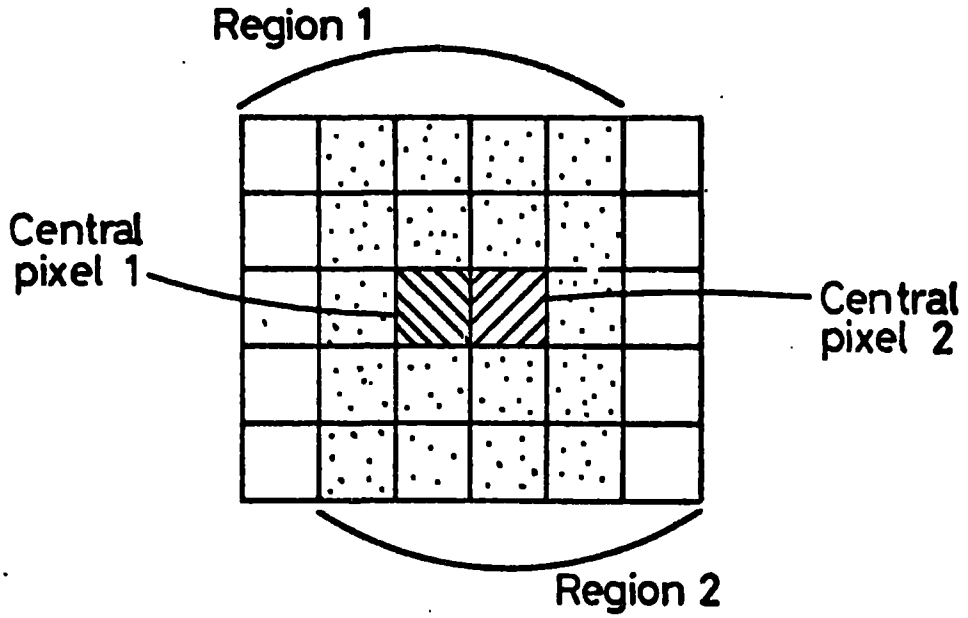


Fig. 7.6 Overlap between adjacent fitting regions

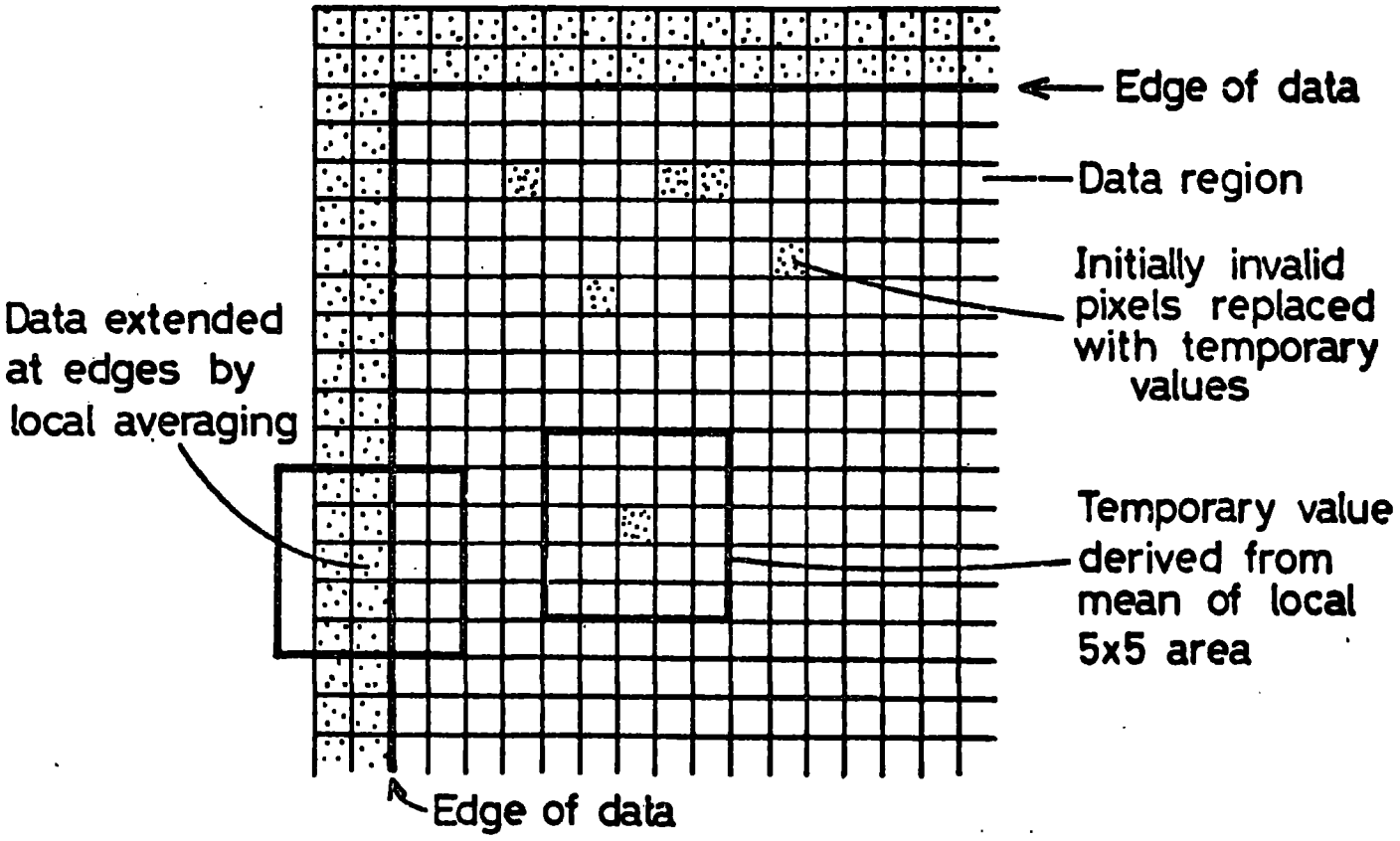


Fig. 7.7 Filling holes in the data prior to fitting

the fitting region.

To overcome this, invalid pixels and regions just outside the data edges are temporarily replaced with smooth "working values" derived by a simple averaging process over the surrounding 5 x 5 pixel area of data. In this way, the fitted function can be evaluated at every valid data point without encountering breaks in the data (Figure 7.7).

7.3.5 Testing the Data for Validity

By using the fitted 3-D function, we can estimate the expected value at every point on the 8 images from the surrounding data and calculate the deviation of the measured value from the expected value. We must now decide when this deviation is acceptable.

Deviations will arise because of random and systematic errors (7.2.3). Considering, for the moment, only the random component, this will cause the deviations to be normally distributed with some standard deviation. Hence we can decide to reject a data point if its deviation from the expected value is more than, say 3 standard deviations. The problem now is to determine the standard deviation expected at a given point in the image.

As equation 7.4 shows, we expect the random noise to be a function of the density on the electronograph, independent of position. (It is the preservation of this positional independence which is the prime reason

for the constant noise interpolation technique of section 5.3.10). Hence we can parameterise the expected standard deviation at every point on the image by some function of the density, and we have a criterion for accepting or rejecting a pixel on the basis of how its deviation compares with that expected.

7.3.6 Noise Against Density

The function describing random noise must be determined empirically, since it will not be the same as the theoretical result (Equation 7.4) to the required accuracy.

This is accomplished by fitting the 3-D function, i_{TH}, to the data at every data point and recording the deviation, δ. The values of δ^2 are then recorded in bins corresponding to the density at that point as determined from the fit (Figure 7.8). This results in a series of estimates of the RMS deviation as a function of density. Eight such sets of bins are used, corresponding to the 8 images. The bins are chosen on a "square root" basis, corresponding to equal increments of the function $\sqrt{\text{density} + 5 \text{ PDS units}}$. This has the advantage of making the resulting function more nearly a straight line so that it can then be easily smoothed by fitting a polynomial to the binned quantities. (The constant 5 PDS units is included so that small negative values of density can be handled). Figure 7.9 shows a typical

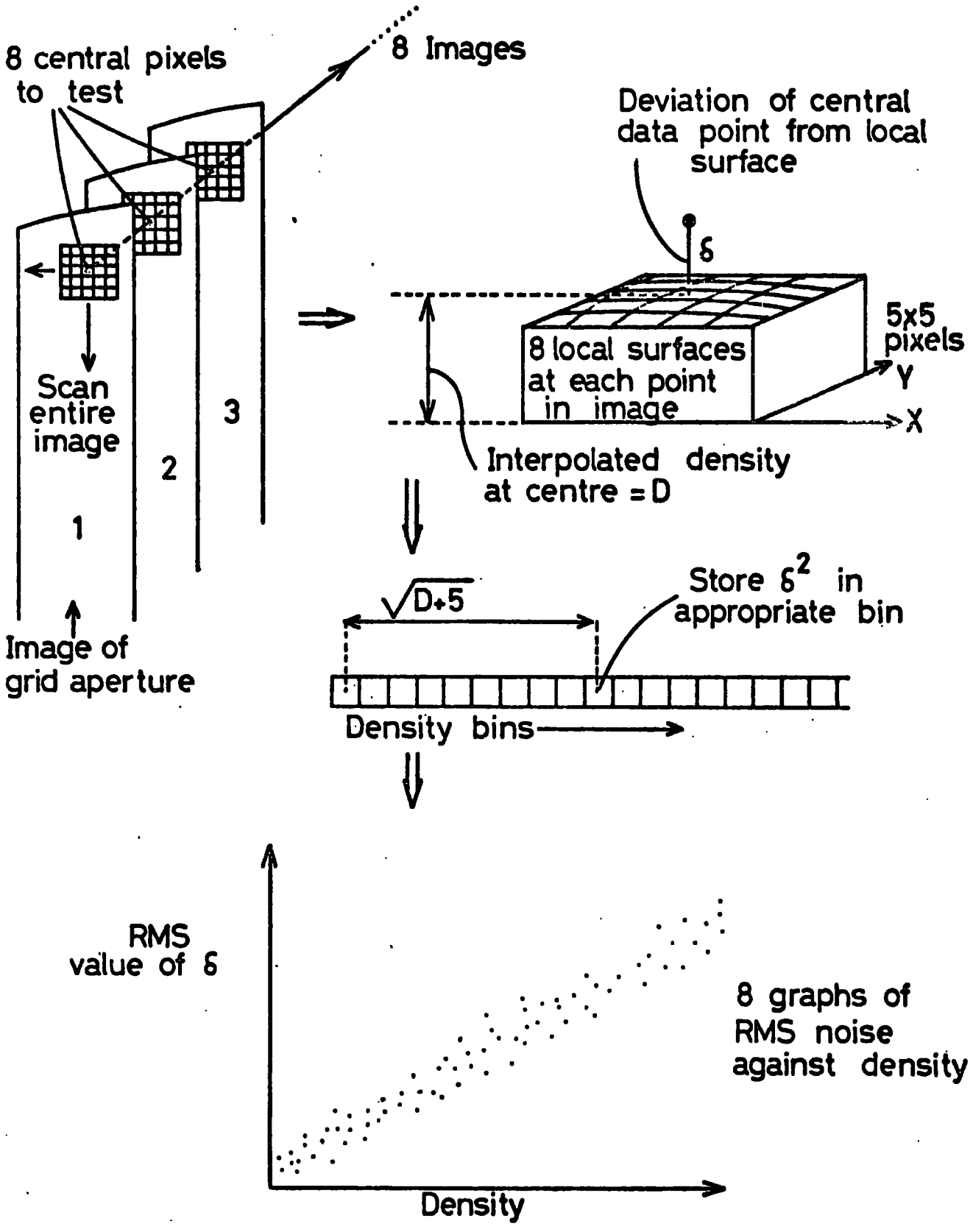


Fig.7.8 Deriving the noise against density functions

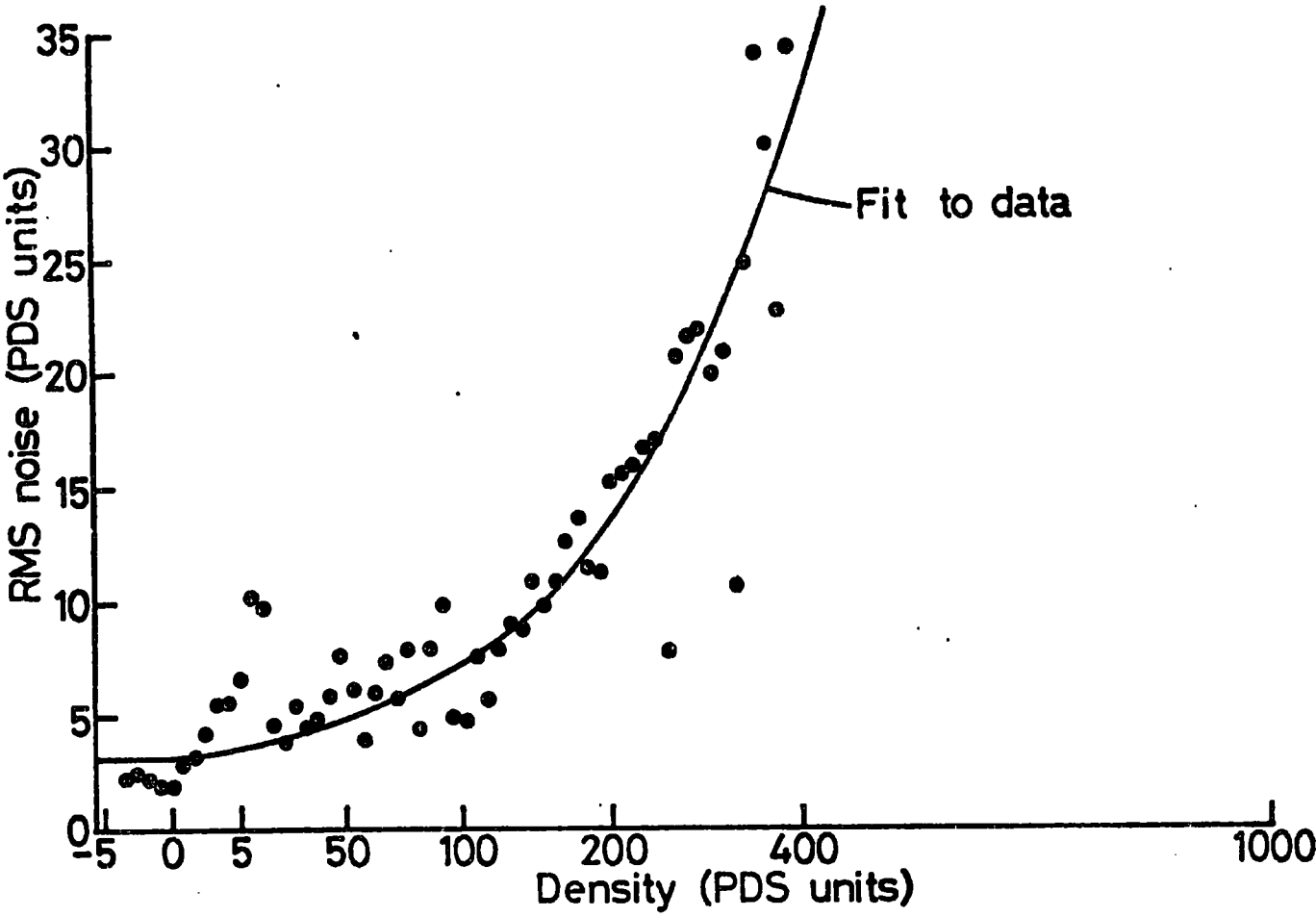


Fig. 7.9 Noise against density for a typical image

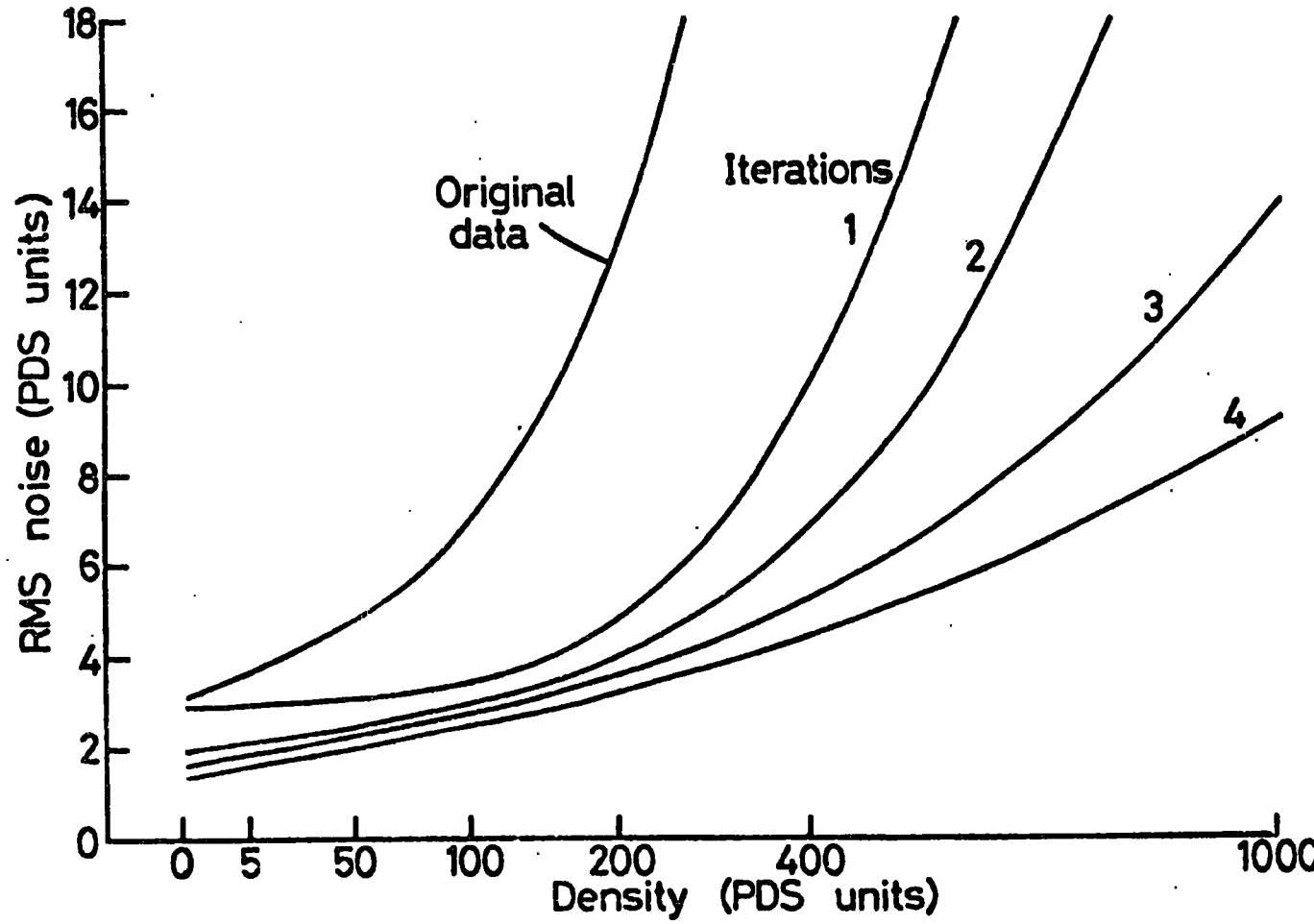


Fig. 7.10 The effect of successive iterations

result obtained from a single image.

7.3.7 Smoothing the Noise Against Density Function

Smoothing of the noise against density function is obviously necessary before it can be used as a basis for rejecting bad data. This is accomplished by fitting a polynomial with the form:

$$P(d) = a_0^2 + a_1^2 d + a_2^2 d^2 + a_3^2 d^3 + a_4^2 d^4 \qquad 7.19$$

where d is the density bin number. This form is used since it reproduces several properties which the noise against density function must have, namely being positive and monotonically increasing for d>0. The parameters a_o to a_4 are determined using a least-squares criterion to minimise.

$$f = \sqrt{\frac{1}{n} \sum_j \left\{ \frac{\sigma_j - P(d_j)}{\sigma_j + 0.2} \right\}^2} \qquad 7.20$$

where n is the total number of data points and σ_j are the RMS deviations within each bin. The least-squares process is weighted such that the fractional square error is minimised, the constant 0.2 being included to prevent excessive weights being generated.

Because the data has not undergone any rejection at this stage, it is necessary to reject some bins here in order to obtain a good fit. Those points which have:

$$\left| \frac{\sigma_j - P(d_j)}{\sigma_j + 0.2} \right| > 2.5\, f \qquad 7.21$$

are rejected, and the fit re-evaluated. Finally the fitted values for densities less than zero are replaced with the value of the fit at zero density. A typical fit obtained by this technique is shown in Figure 7.9.

7.3.8 Iterations

As figure 7.9 shows, the experimental estimates of noise have considerable errors due to the presence of corrupt data, and even the smoothed curve has considerably higher values than theory would predict (Equation 7.4). This function, therefore, can only be used as an approximation, and a more accurate one must be derived after the most corrupt data points have been detected and removed. We thus arrive naturally at the idea of successive iterations involving more thorough rejection of bad data each time. This also has a number of other advantages:

i) Because some corrupt data can be very bad (e.g. a value of 1 can be corrupted to 1000 by the presence of a speck of dust on an electronograph) the 3-D function fitted to the data can be considerably distorted by a single bad pixel and this could lead to the erroneous rejection of neighbouring pixels. It is thus desirable

to reject the worst data first, returning later to the more doubtful cases.

ii) When a pixel is rejected, it can be replaced temporarily at each iteration with the value interpolated from the local 3-D fit. After several iterations these replacement values will become good interpolations and will not affect the fit to neighbouring points.

The overall procedure adopted is thus:

i) Replace initially invalid pixels with a local mean "working" value, and extend the edges of the data by the same means.

ii) Fit the 3-D function to the entire set of data, deriving 8 curves of noise against density from the deviations encountered, rejecting those pixels whose deviations are very large and replacing all invalid points with the value interpolated from the 3-D fit.

iii) Fit the 3-D function to the entire set of data, rejecting those pixels whose deviations are more than g x the standard deviation expected from their density (initially g = 8). Form 8 curves of noise against density from those pixels which are accepted and replace all the invalid pixels by their new interpolated values.

iv) Repeat step (iii) for g = 4, 3, 2.5, 2, 2 ...., the number of iterations depending on the quality of the data and the amount of rejection required. Successive estimates of noise against density obtained by this means are shown in figure 7.10.

7.3.9 Systematic Errors

So far we have neglected the systematic errors present on each image. The effect of these errors is shown in figure 7.11, which shows the deviations from the fitted 3-D function within one image, plotted against the position in the image. The ideal curve represents random noise, but in practice there are large scale variations due to the systematic errors. Since these errors will affect our estimates of noise against density and also our ability to detect invalid data amongst the random noise, some method must be found of separating the two sources of error.

The solution adopted is to store the deviations from the 3-D fit in a matrix after each iteration. These deviations are then smoothed in 2 dimensions by replacing each value by the mean of the mxm area of pixels centred on it (the value of m is discussed below). The result, referred to as the "smoothed residuals" represent the systematic errors, since the random errors are cancelled by the smoothing process. These smoothed residuals are then subtracted from the data before the next iteration, such that the deviations encountered in the next iteration are due to random errors alone. The nett effect of this process is that the systematic errors are removed to a better and better approximation as the iterations proceed and the data becomes more and more pure.

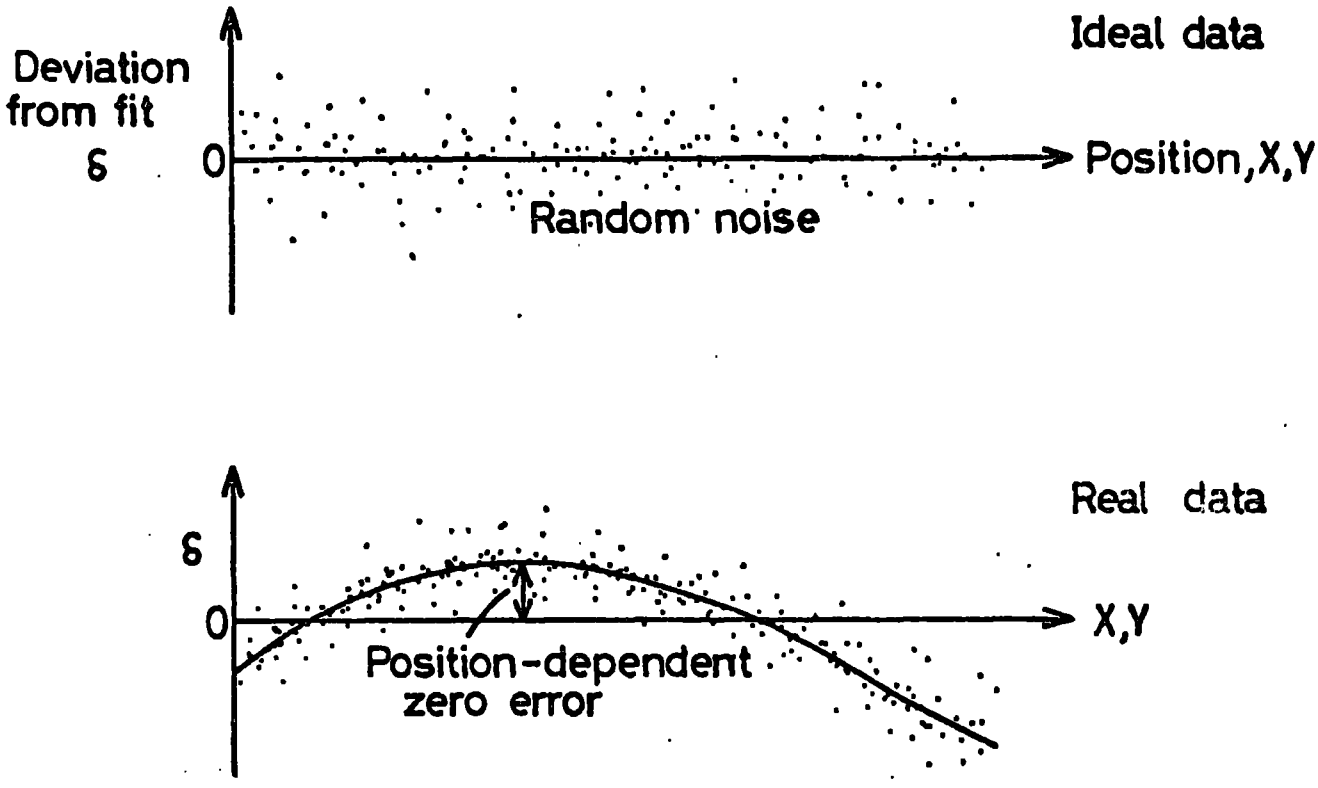


Fig. 7.11 Effect of systemmatic errors on the residuals

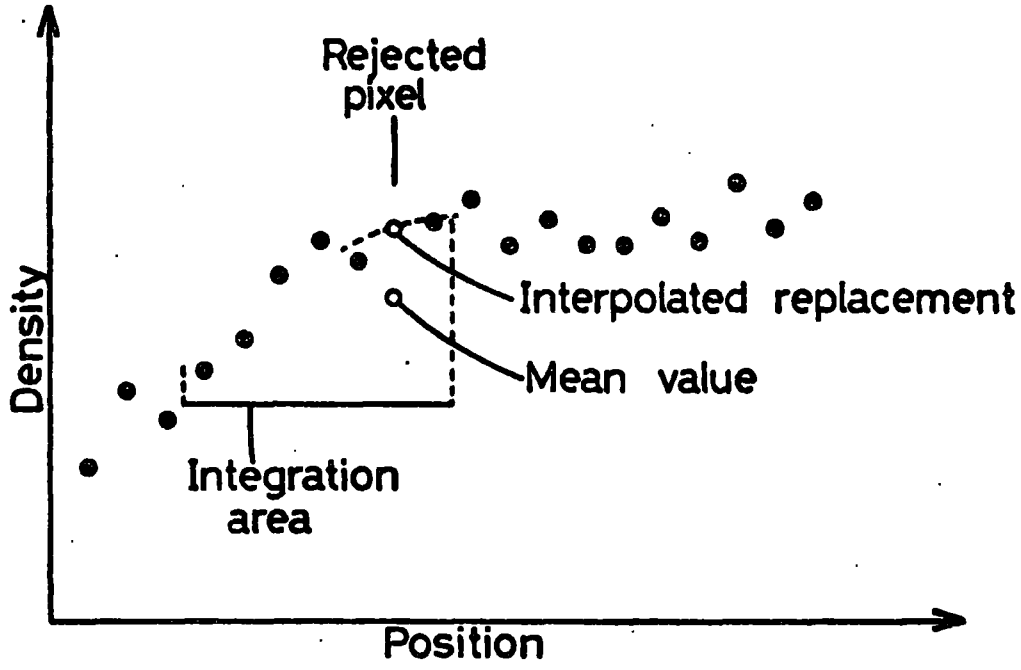


Fig. 7.12 The use of interpolated replacement pixels

7.3.10 Image Independence

It can be shown that if there are large scale systematic errors e_1 to e_8 on the 8 images, the corresponding smoothed residuals R_1 to R_8 are given by the matrix equation:

$$(R_1,R_2,R_3,R_4,R_5,R_6,R_7,R_8) = \frac{1}{8}\begin{bmatrix} 5 & 1 & -1 & -1 & 1 & -3 & -1 & -1 \\ 1 & 5 & -1 & -1 & -3 & 1 & -1 & -1 \\ -1 & -1 & 5 & 1 & -1 & -1 & 1 & -3 \\ -1 & -1 & 1 & 5 & -1 & -1 & -3 & 1 \\ 1 & -3 & -1 & -1 & 5 & 1 & -1 & -1 \\ -3 & 1 & -1 & -1 & 1 & 5 & -1 & -1 \\ -1 & -1 & 1 & -3 & -1 & -1 & 5 & 1 \\ -1 & -1 & -3 & 1 & -1 & -1 & 1 & 5 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \end{bmatrix}$$

7.22

As can be seen, there is a large amount of coupling between some pairs of images, such that it is difficult to tell from the residuals which image contains the error. Because of this, it is quite possible to reject data in a perfectly good region of image because of a particularly bad region on another image. The smoothed residuals provide a means of adjusting this inter-dependence of the images as follows:

If the deviations are smoothed over a small area- such as the 5 x 5 area over which the 3-D function is fitted - this has the effect of making the fits to

different images largely independent as regards systematic errors, because each individual fitted region of 5 x 5 pixels can then have a constant added to it on each of the 8 images independently of its surroundings. Data rejection in this case relies on the smoothness requirement within each image, rather than on the consistency between images.

If the region over which the deviations are smoothed is made large, then local deviations on one image cannot be followed by the smoothed residuals and will not be subtracted unless they are of large extent, and the image interdependence is fully used to check for internal consistency between the 8 images on all scales up to that over which the residuals are smoothed.

In the light of the above, the smoothing region has been made small in the initial iterations, where very bad data points are present and maximum image independence is required, but becomes progressively larger as the iterations proceed to take advantage of the more accurate estimates of noise which are derived, and to enable those regions to be rejected which are smooth on one image although they do not agree with the other 7 images. The actual values used (m in the previous section) are 5 x 5, 7 x 7, 9 x 9, 11 x 11, 13 x 13 in successive iterations.

7.3.11 Output Data Form

When the final iteration is reached in the data validation process, maximum use is made of the information obtained to enable the next stage of data reduction to evaluate the errors on the result.

Two sets of data are produced; the first contains all the data points which have not been rejected, together with interpolated values for those which have been rejected (as derived from the 3-D fit in the last iteration), and the second set contains the deviations from the 3-D fit for each valid pixel (from the last iteration) together with flags to indicate which pixels are interpolations rather than true data points.

The value of using the interpolated value for rejected pixels, rather than simply omitting them from the dataset, is illustrated in figure 7.12. Since the pixels must be grouped together to reduce the noise in the final result, omitting a pixel amounts to replacing it with the mean value within its group. As figure 7.12 shows, this can be considerably less accurate than an interpolated value.

The overall data validation technique is summarised in figure 7.13.

7.4 Practical Results of Data Validation

7.4.1 Rejection of Bad Data

The overall effect of the data validation process is shown in figure 7.14, which shows a polarimeter image,

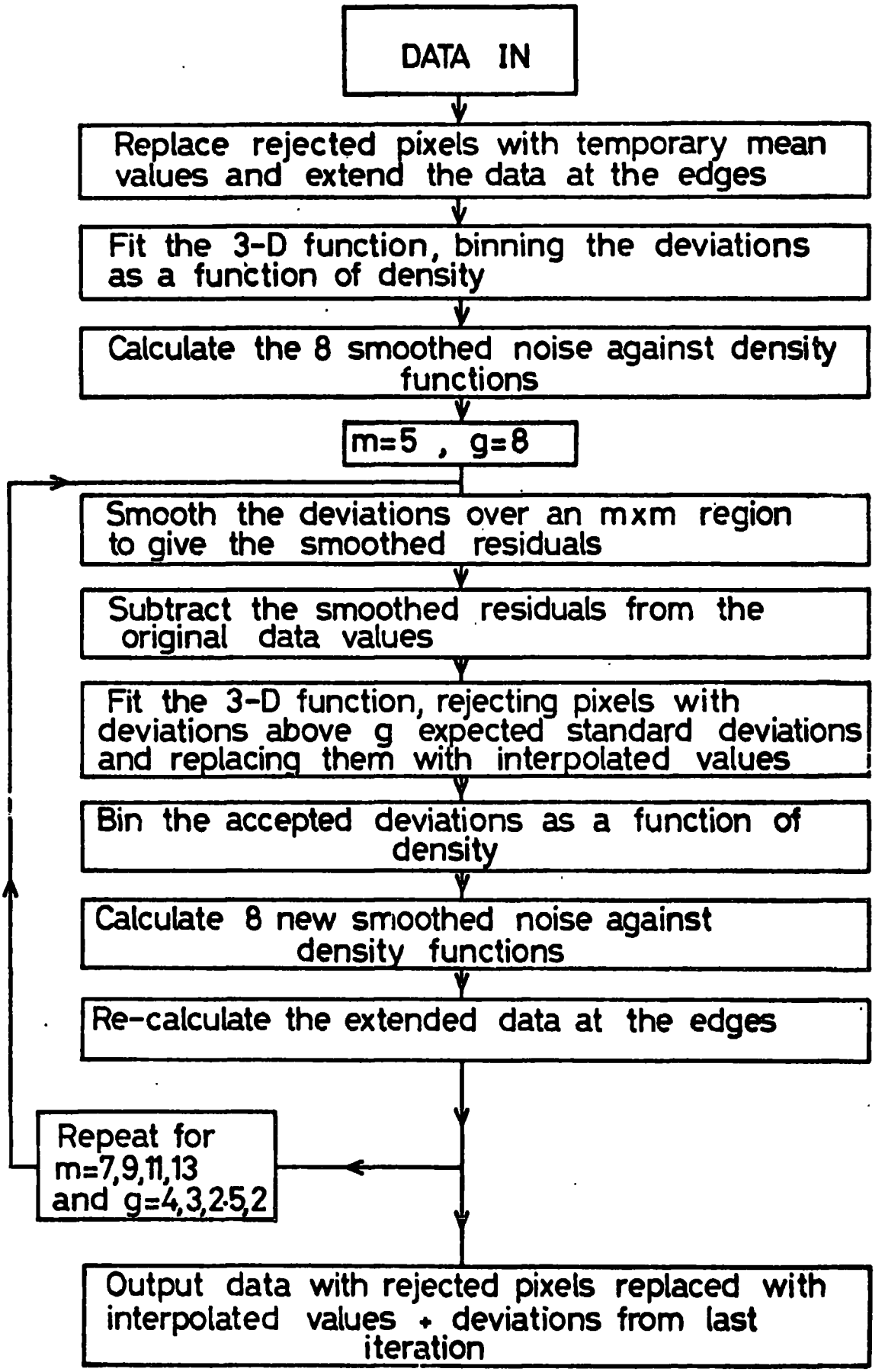


Fig. 7·13 Summary of the data validation process

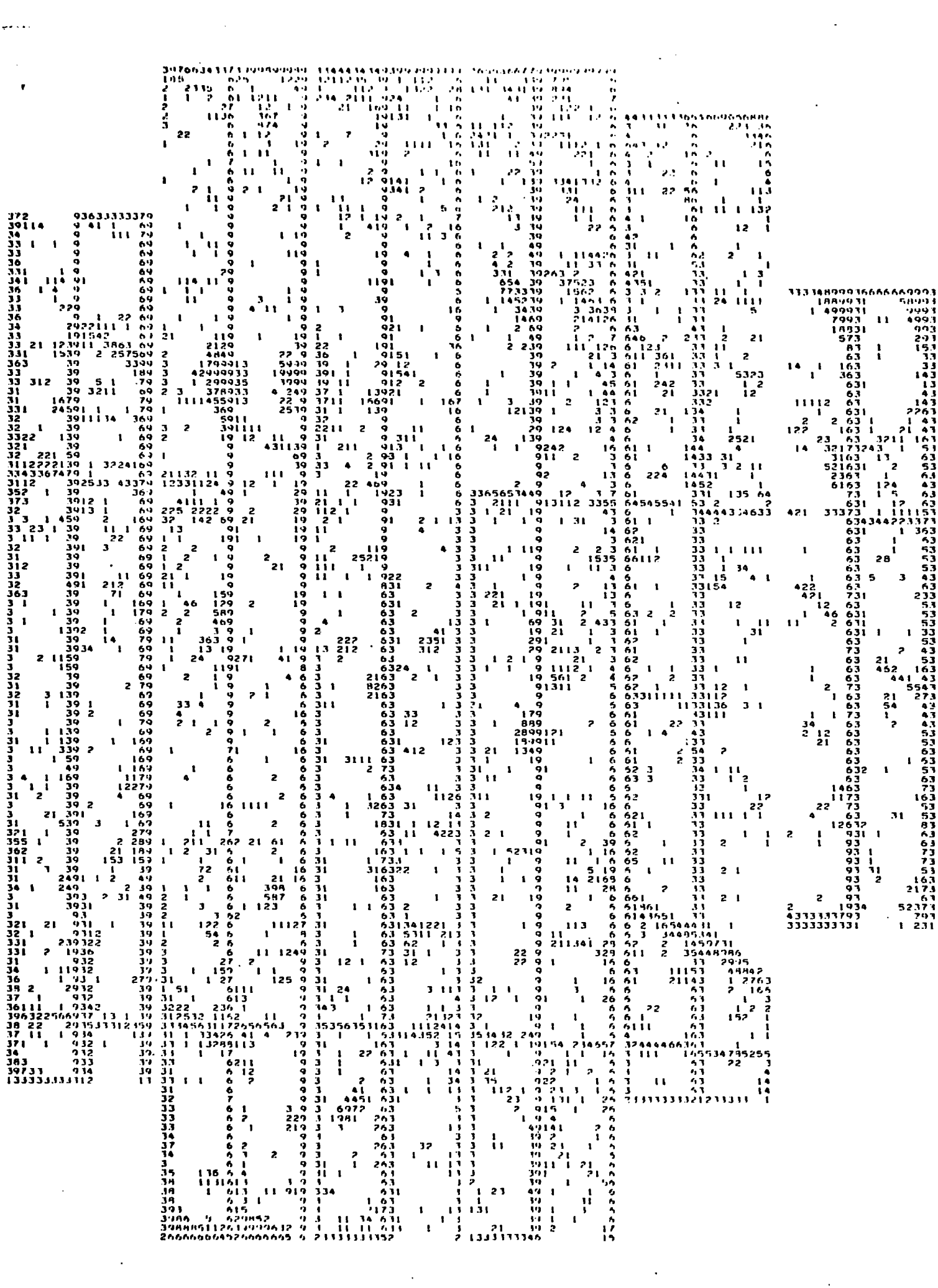


Fig.7.14 Pixels rejected in a polarimeter image

indicating the number of pixels rejected at each location. This number runs from 0 (blank) to 9 (each character represents a 3 x 3 pixel region of data) and the vertical columns of numbers are caused by the grid-overlap rejection at the edges of the images, rejected during photocathode correction (5.5.8).

Although the figure shows a large number of regions containing rejected pixels, the actual number rejected is less than 5% and as can be seen, major blemishes such as dirt blobs and scratches account for most of these.

Figure 7.15 shows a similar result, but now comparing the 8 images of the same grid aperture (each character here represents a 2 x 2 pixel region of data). As can be seen, there is little correlation between the regions rejected on each image; scratches on one image do not cause rejected pixels on other images.

A more detailed examination of the rejection process is shown in figure 7.16 which shows traces of the intensities i_1 and i_2 both before and after data validation, showing the interpolations which have replaced the rejected pixels.

An example of the effect on the final polarisation map will be found in section 9.6.3.

7.4.2 Random Errors (Experimental)

The data validation process provides an opportunity to examine the practical errors in electronography in

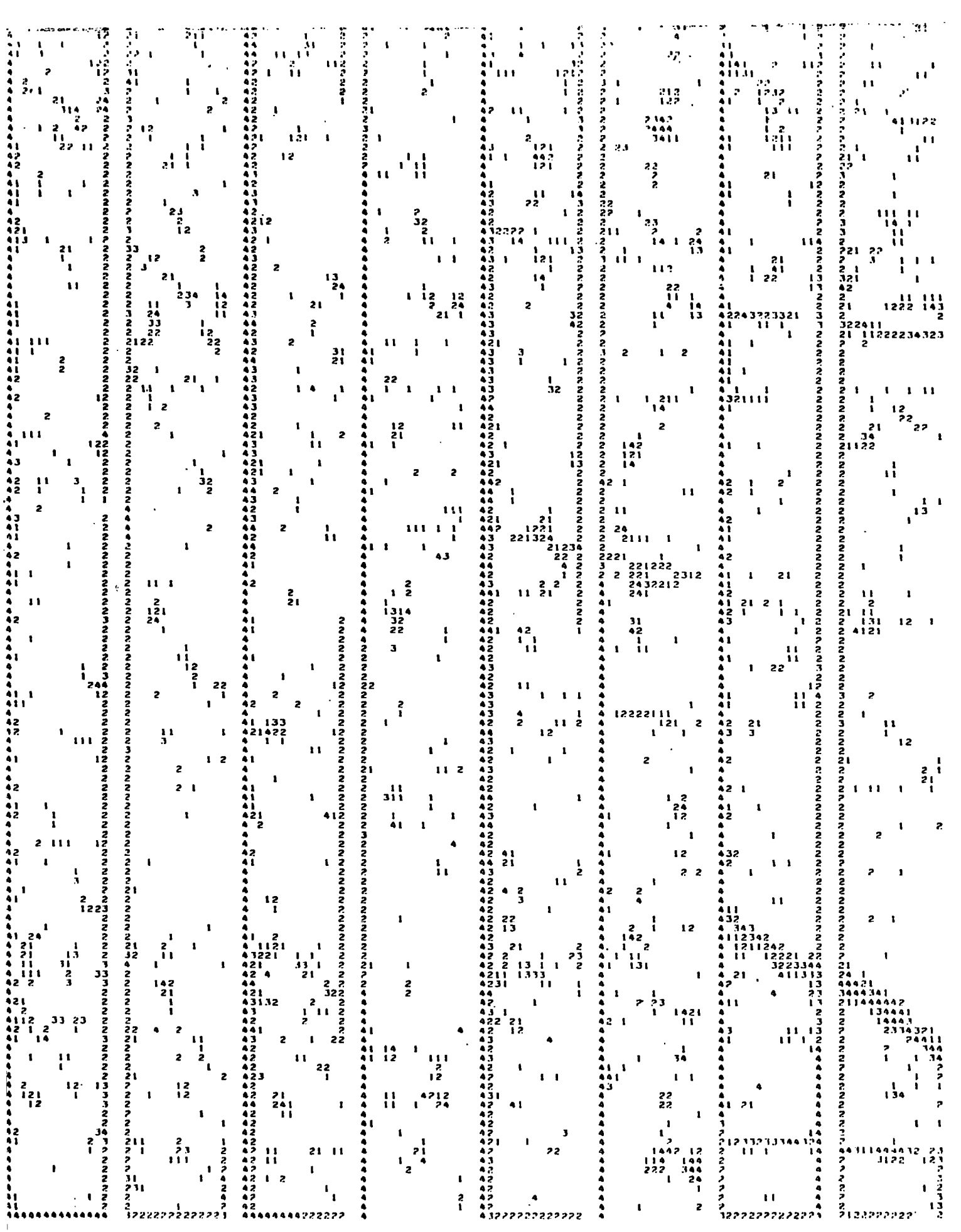


Fig.7.15 Pixels rejected in the 8 images of a single grid aperture

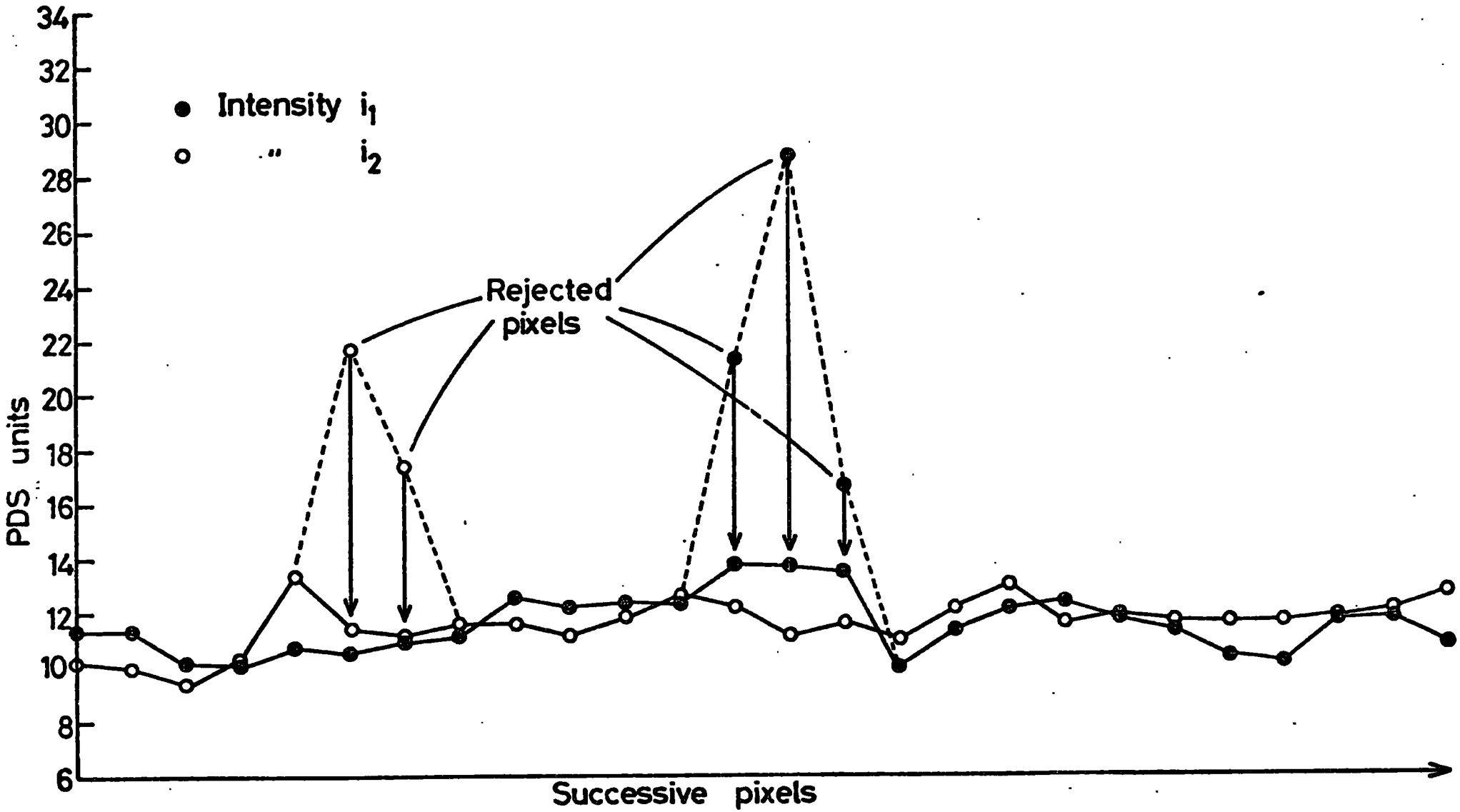


Fig.7-16 Detection and replacement of bad pixels

some detail. We will return to this in a later chapter, but the practical results are presented here.

Figure 7.10 shows the estimates of random noise against density for successive iterations (7.3.8). As can be seen, the initial estimate is far from accurate and later estimates show a considerable reduction of noise, indicating more sensitive data rejection.

The comparison with theory is also of interest. Figure 7.17 shows the experimental values of noise against density for the final iteration on a reasonably noise-free set of data, plotted alongside the theoretical curve (equation 7.4), allowing for the value of sky background subtracted, the scale factor applied, and the reduction in noise by a factor of 2 due to the constant noise interpolation technique (5.3.10).

As can be seen, the experimentally derived noise estimates are in close agreement with theory, indicating that photon noise accounts for practically all the random noise on the result. The wider dispersion of the experimental values at high density is due simply to the smaller number of pixels at high density from which to measure the noise.

There are, however, two factors which could invalidate this comparison. Firstly, the measurements of each pixel may not be independent, due to the time constant of the PDS measuring machine. If this were the case, a lower estimate of the noise would result.

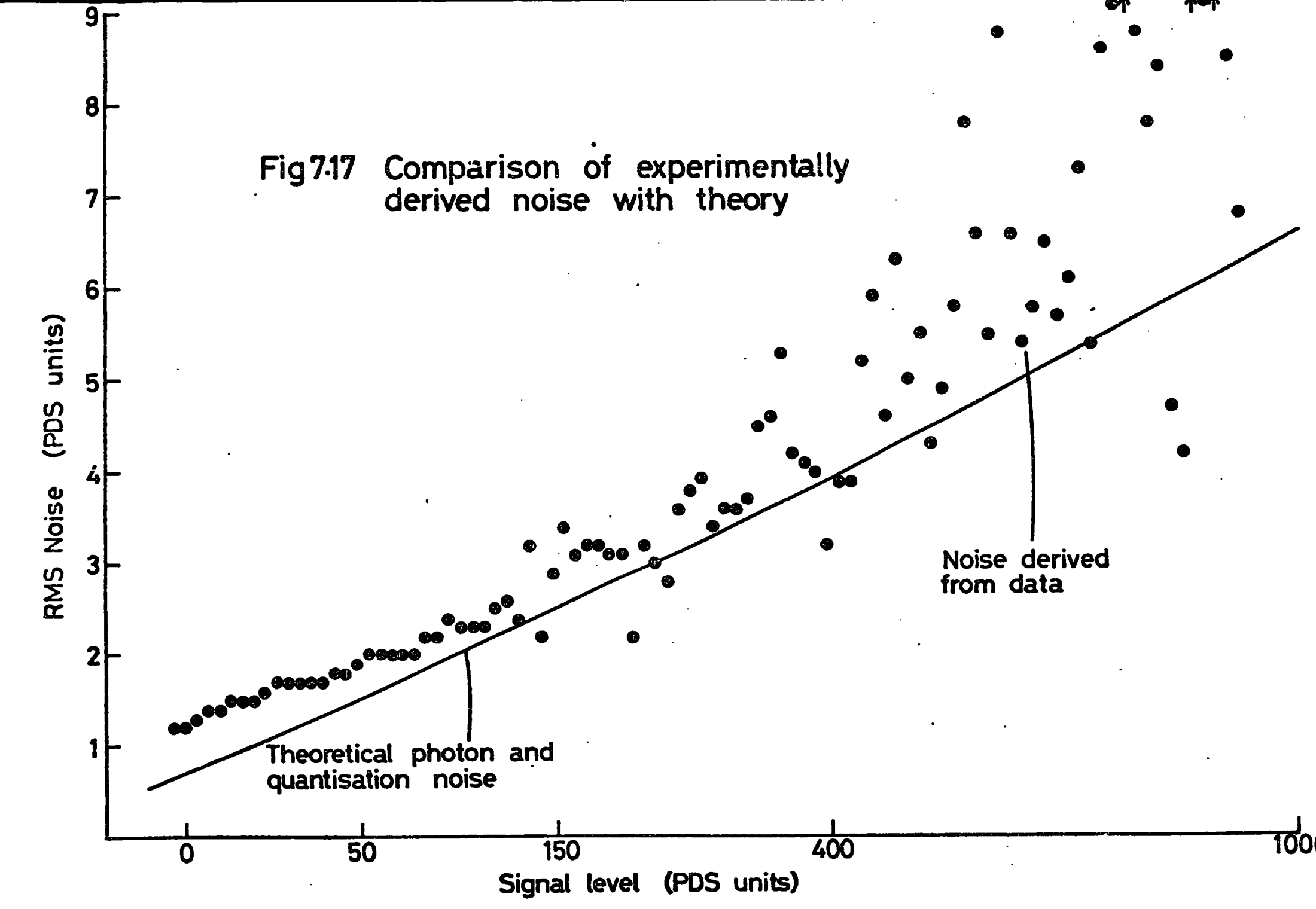


Fig 7.17 Comparison of experimentally derived noise with theory

Such a condition is hard to justify, however, since it would also lead to blurring of small image features (e.g. dust specks) which is not observed. A possible test also, is to determine whether the measured noise depends on the speed at which the PDS machine is scanned. For the limited number of cases where this has been done, no difference was found.

Secondly, the comparison rests on the value of film sensitivity used to calculate the theoretical curve. There is, however, evidence that this may be variable. In this case, an accurate calibration would be required for this comparison to be valid.

In spite of the above comments, it seems justifiable to believe that there are no major sources of random error in the system apart from photon noise.

7.4.3 Systematic Errors (Experimental)

The systematic errors can be studied via the smoothed residuals found during data validation (7.3.9). Figure 7.18 shows typical residuals smoothed over a region of 11 x 11 pixels, (each character represents a 2 x 2 pixel region) and presented for all 8 images within a grid aperture. The scale 1,2,3,4,5,6,7,8,9 corresponds to residuals of -4,-3,-2,-1,0,1,2,3,4 PDS units, with +,- indicating out of range. As can be seen, the residuals are usually within 1 or 2 units of zero. There is some correlation with the image content of the data, indicating

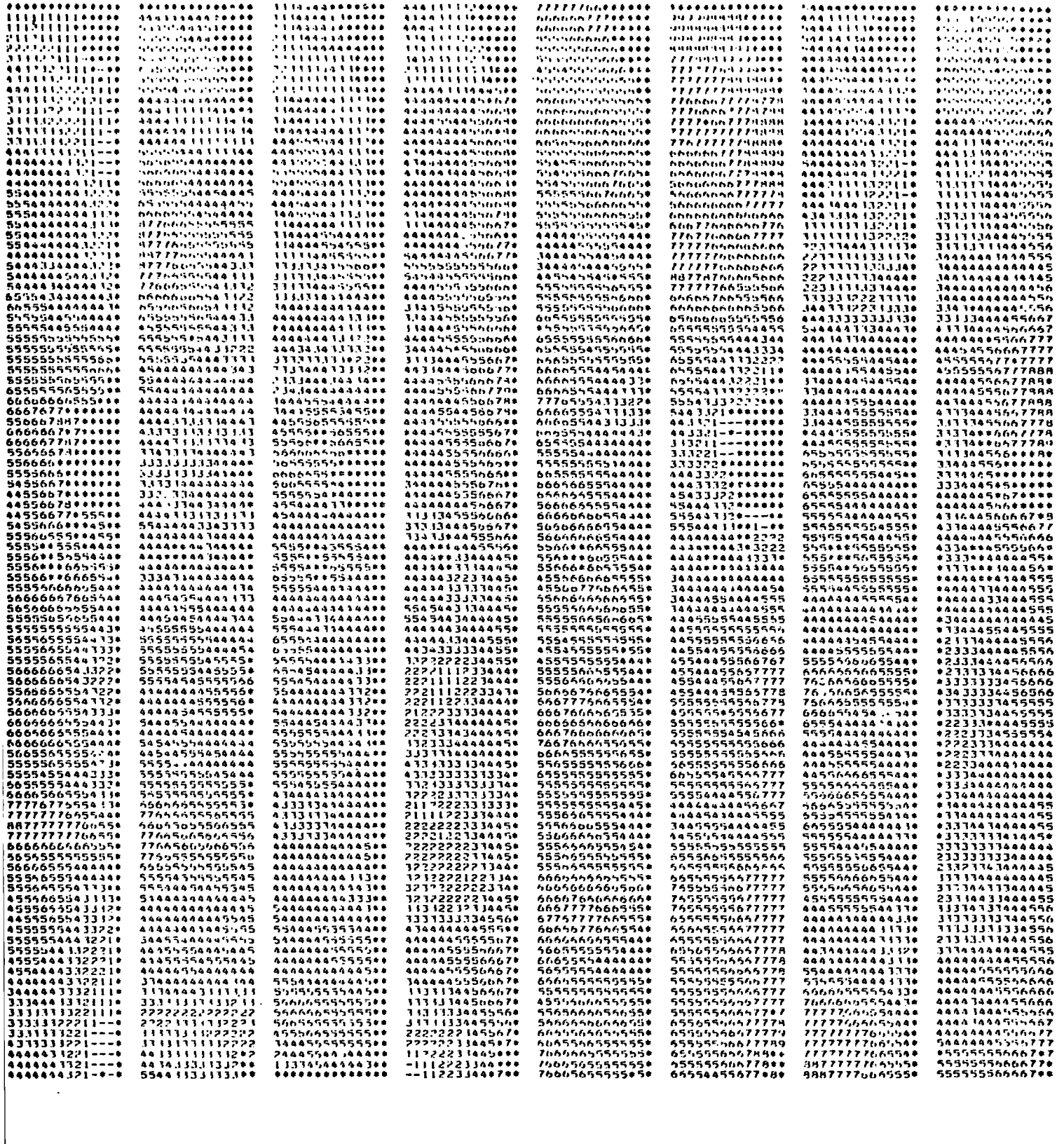


Fig.7.18 Typical smoothed residuals on a set of 8 images

that the linearity corrections or scale factors applied are not perfect (although this is a very sensitive test and the errors are quite acceptable).

We also have a good check on the sky subtraction at this stage, since an error in zero level on one image will lead to an overall deviation of its smoothed residuals from zero, according to equation 7.22.

7.4.4 Limitations of the Technique

While the technique outlined has many advantages and has proved very versatile, being completely automatic when applied to a wide range of data, there are a number of limitations which must be born in mind when using it. Since data reduction can proceed without validation, this course can be adopted if there is any doubt about its applicability. The main limitations are:

i) If the star images are small, due to good atmospheric seeing and/or the image scale of the telescope, then the 3-D function used does not give a good description of the star images. This problem is aggravated by very small degrees of mis-alignment between the images which defeat the 3-D function which tries to fit them all simultaneously.

The prime effect of this - that star images are rejected as though they were scratches on the electronographs - is of no consequence, since the polarimeter is not designed for stellar measurements and no meaningful measurements can be made in the vicinity of bright

stars. The important effect is that this leads to an over-estimate of the noise at typical stellar densities, resulting in inefficient data validation in other regions of similar density. Recently an attempt has been made, with some success, to overcome this problem by specifying that some regions of the image are not used when calculating the noise.

ii) A similar effect to (i) above is produced if the images are made mutually inconsistent due to changing atmospheric seeing during the observations. This primarily affects the star images.

iii) If regions of the electronographs are over-exposed, effects can be caused at the edge of the over-exposed region due to the sudden lack of data. This only becomes important if measurements are needed as close as possible to the over-exposed region. The excessive rejection of data which can occur can be avoided by using such regions without any data validation.

7.5 Calculation of the Polarisation Parameters

7.5.1 Binning the Intensities

The first step in converting the 8 images into a polarisation map is to break the images into integration areas and average the intensity within these areas to reduce the statistical noise on the result. These areas may be square or rectangular and contain, typically, between 9 and 100 pixels.

Within an integration area, each image may have any number of pixels which are invalid for some reason, and because we need reliable estimates of the integrated intensities, we may have to reject those where the number of valid pixels is too low. In general, therefore, we have to calculate the polarisation parameters from a set of 8 or less integrated intensities. To decide whether to accept an integrated intensity, or not, the following procedure is employed:

The total number of pixels within the integration area on each image (valid and invalid pixels) is scanned to find the maximum value. This may be less than the area of integration if the area is centred near the edge of the image (Figure 7.19), but if it is less than 50% of this area, no measurement is made at that point. The number of valid pixels within each integrated intensity is then found, and if it is less than 80% of the maximum number of pixels, that intensity is rejected. In this way, only those intensities, where valid pixels fill 80% of the image area inside the integration area, are used.

7.5.2 Forming the Integrated Intensities

Having decided which integrated intensities to accept, the mean intensity within the integration area on each image is found from:

$$\int = \frac{\Sigma i}{n_T} \qquad 7.23$$

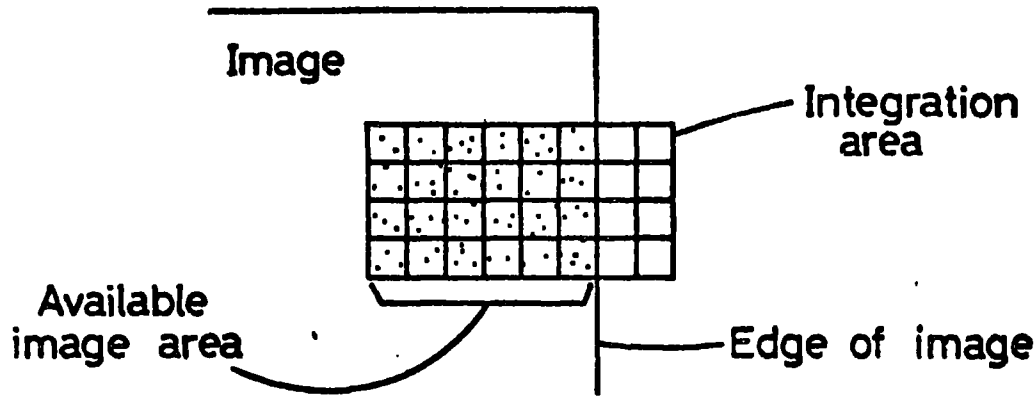


Fig.7.19 Integration area near edge of image

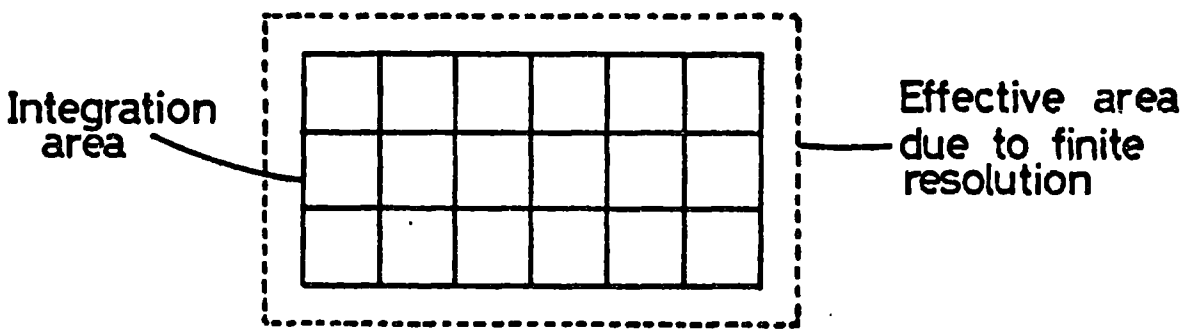


Fig.7.20 Correction due to finite resolution

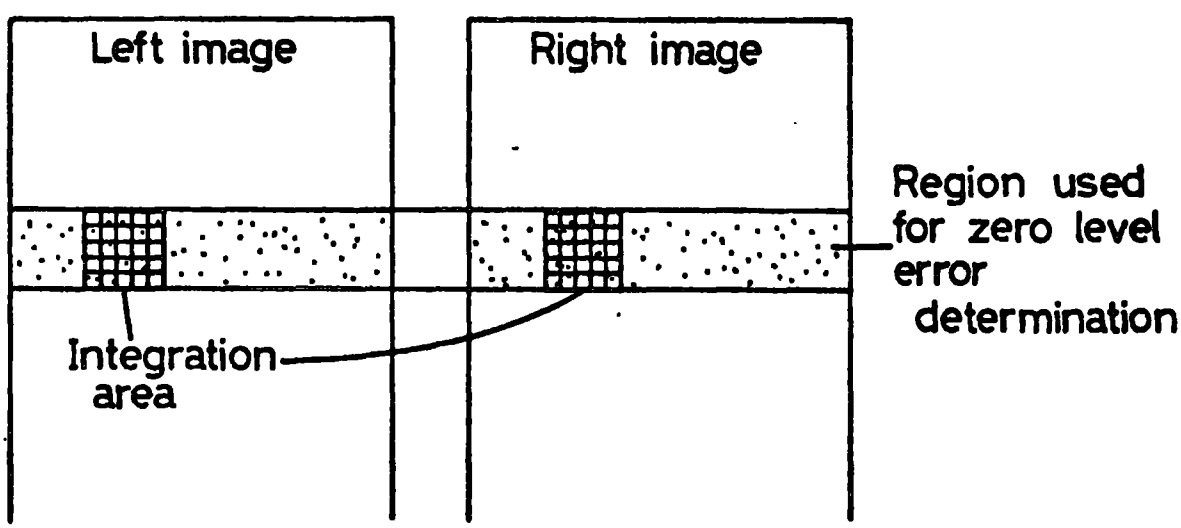


Fig.7.21 Averaging the deviations to give the zero level errors

where the summation is over all the pixels (valid data and interpolated values), and n_p is the total number of pixels.

We also calculate an estimate of the errors on ϑ from the deviations from the 3-D fit produced during the data validation process. In the presence of systematic errors, the mean deviation within an integration area is not necessarily zero (7.3.9), so the mean value is found from:

$$R = \frac{\Sigma \delta}{n_g} \qquad 7.24$$

where the summation is over all the deviations for valid pixels, and n_g is the number of these valid pixels.

The random error is estimated from the dispersion of the deviations about this mean value:

$$\sigma = \frac{2}{0.7737} \sqrt{\frac{nm}{(n+1)(m+1)}} \sqrt{\frac{\Sigma(\delta-R)^2}{n_g^2}} \qquad 7.25$$

The factor 2 allows for the smoothing introduced by the constant-noise interpolation (5.3.10), while the factor 0.7737 is included because deviations greater than 2σ from the mean have already been rejected during data validation and omission of these points gives too small an estimate of σ. The further correction factor

used is to account for the fact that some image smoothing has taken place during image transformation, so that each integration area contains information from an area about ½ pixel wide around its edge (The integration area is m x n pixels), hence the error calculated is based on an integration area of (m + 1) x (n +1) pixels. This correction is only important when the integration area is small (Figure 7.20).

7.5.3 Handling the Systematic Errors

As has already been pointed out, the systematic errors can be divided into two types: those which vary on a scale large compared with the image separation on the electronographs and thus affect both images together, and those with smaller scale variations which affect the difference between the two images.

The large scale errors affect the measurement of total light intensity, but not the measurement of polarisation, while the small scale errors affect primarily the polarisation measurements. We thus need to remove the large scale variations before we can calculate the effect on the polarisation produced by the smaller-scale variations.

The large scale variations (referred to as zero level errors) are found by integrating the deviations from the 3-D fit over a large area of the electronograph containing the measurement area (Figure 7.21). To prevent errors due to bright star images within this area

deviations corresponding to densities above 500 PDS units are not included. The mean zero level error is found in this way for each of the 4 electronographs, and is subtracted from the integrated intensities (Equation 7.23) and the mean deviations (Equation 7.24) to leave only the small scale systematic effects. The effect of the zero level error itself is included in the total intensity evaluation at a later stage (7.5.7).

7.5.4 Weighting the Intensities

When calculating the polarisation parameters, we will obtain the most accurate result if we give more weight to those intensities with least probable error. In some cases, random errors will dominate and we can derive weights from these. In some cases, however, systematic effects will dominate and we then want to give all intensities equal weight, since accuracy then depends on averaging out the systematic errors on as many images as possible.

To achieve this discrimination, we must decide whether the observed mean residuals R_k (k=1,8) after zero level error subtraction, can be explained entirely in terms of random errors, or whether small scale systematic errors must play a part. We can then compare the relative magnitudes of these effects and derive weights accordingly.

If each of the integrated intensities J_k has a

probable RMS error ε_k, due to random and systematic components σ_k and s (assuming the same RMS systematic error on each image at a particular point), where:

$$\varepsilon_k^2 = \sigma_k^2 + s^2 \qquad 7.26$$

Then equation 7.22 can be used to show that the probable RMS value of the mean residuals R_k is given approximately by:

$$\sqrt{\widehat{R^2}} = \sqrt{\frac{5}{8} \cdot \frac{\Sigma \varepsilon_k^2}{N}} \qquad 7.27$$

where N is the number of integrated intensities. (Having removed the large scale systematic errors, the errors on the individual intensities are now independent).

Hence we can estimate $\Sigma\varepsilon_k^2$ from:

$$\Sigma \varepsilon_k^2 = \frac{8}{5} \Sigma R_k^2 \qquad 7.28$$

and hence estimate:

$$s^2 = \frac{1}{N} \left[\frac{8}{5} \Sigma R_k^2 - \Sigma \sigma_k^2 \right] \qquad 7.29$$

If $s^2 < 0$, no systematic errors are necessary to explain the values of R obtained, random errors being sufficient, hence we set s = 0, otherwise we use the value of s^2 obtained to derive weights for each integrated intensity:

$$w_k = \frac{1}{\sigma_k^2 + s^2} \qquad 7.30$$

This procedure achieves the aims outlined at the beginning of this section, allowing the weights to adjust to the cases of dominant random or systematic errors automatically.

7.5.5 The Least-Squares Fit

Having determined up to 8 integrated intensities, with corresponding weight factors based on their probable total error, we can perform a least-squares fit using the known form of the polarisation data (Equation 7.6) to obtain the most probable values of the Stokes parameters. We need to minimise:

$$f = \sum_k \left[g_k - (I + Q\cos 2\theta_k + U \sin 2\theta_k) \right]^2 w_k \qquad 7.31$$

with respect to I,Q,U, where summation is performed over all the integrated intensities present. Differentiating and equating to zero yields 3 normal equations:

$$\Sigma w_k + Q\Sigma \cos 2\theta_k w_k + U\Sigma \sin 2\theta_k w_k = \Sigma g_k w_k$$

$$\Sigma \cos 2\theta_k w_k + Q\Sigma \cos^2 2\theta_k w_k + U\Sigma \sin 2\theta_k \cos 2\theta_k w_k = \Sigma g_k \cos 2\theta_k w_k$$

$$\Sigma \sin 2\theta_k w_k + Q\Sigma \cos 2\theta_k \sin 2\theta_k w_k + U\Sigma \sin^2 2\theta_k w_k = \Sigma g_k \sin 2\theta_k w_k$$

7.32 - 7.34

Setting: $A = \Sigma w_k \cdot \sin 2\theta_k$ 7.35

$B = \Sigma w_k \cdot \cos 2\theta_k$ 7.36

$C = \Sigma w_k \cdot \sin^2 2\theta_k$ 7.37

$D = \Sigma w_k \cdot \cos^2 2\theta_k$ 7.38

$W = \Sigma w_k$ 7.39

$X = \Sigma \vartheta_k \cdot w_k$ 7.40

$Y = \Sigma \vartheta_k \cdot w_k \cdot \sin 2\theta_k$ 7.41

$Z = \Sigma \vartheta_k \cdot w_k \cdot \cos 2\theta_k$ 7.42

and solving for I,Q,U gives:

$$I = \frac{(XDC - YDA - ZBC)}{T} \qquad 7.43$$

$$Q = \frac{Z - IB}{D} \qquad 7.44$$

$$U = \frac{Y - IA}{C} \qquad 7.45$$

where $T = (WDC - DA^2 - CB^2)$ 7.46

The trigonometric functions involved in these calculations are summarised in figure 7.5. Some combinations of integrated intensities cannot be used to find the

polarisation parameters: such combinations are easily recognised, however, since they always give T = 0.

7.5.6 Calculation of Polarisation Errors

The propagation of the errors on the individual intensities through the least-squares fitting process to estimate the errors on the Stokes parameters is performed by differentiating the Stokes parameters with respect to the intensities. Since the errors on the intensities are independent, the probable error on I is given by:

$$\Delta I^2 = \sum_k \left(\frac{\partial I}{\partial g_k}\right)^2 \quad (\sigma_k^2 + s^2) \qquad 7.47$$

with the notation of the last section, this gives:

$$\Delta I^2 = \sum_k \left[\frac{\partial I}{\partial X} \cdot \frac{\partial X}{\partial g_k} + \frac{\partial I}{\partial Y} \cdot \frac{\partial Y}{\partial g_k} + \frac{\partial I}{\partial Z} \cdot \frac{\partial Z}{\partial g_k}\right]^2 \cdot \frac{1}{W_k^2} \qquad 7.48$$

$$\therefore \quad \Delta I^2 = \sum_k \left[\frac{DC}{T} - \frac{DA}{T} \sin 2\theta_k - \frac{BC}{T} \cos 2\theta_k\right]^2 \cdot \frac{1}{w_k^2} \qquad 7.49$$

By a similar process, the errors on Q and U are found.

7.5.7 The Intensity Errors

Before proceding further, we must take account of two other sources of error which do not affect the polarisation parameters but do affect the measurement of total intensity, namely the RMS linearity errors ε_{LIN} (6.4.4) and the RMS zero level errors ε_z (7.5.3).

The 4 zero level errors z_j ($j = 1,4$) can be used to estimate ε_z from:

$$\varepsilon_z = \sqrt{\frac{\sum_j z_j}{12}} \qquad 7.50$$

by using the same argument as for ε_{LIN} (6.4.4).

Combining these sources of error into the estimate of ΔI (Equation 7.49) gives:

$$\Delta I^2 = \sum_k \left[\frac{DC}{T} - \frac{DA}{T}\sin 2\theta_k - \frac{BC}{T}\cos 2\theta_k\right]^2 \frac{1}{w_k^2} + \varepsilon_{LIN}^2 + \varepsilon_z^2 \qquad 7.51$$

The linearity errors have been stored previously and are recalled from a table according to the total intensity measured.

7.5.8 The Polarisation Measurements and Errors

The polarisation parameters commonly required, the polarised intensity I_p, % polarisation P and orientation θ can be derived from the Stokes parameters

as follows:

$$I_p = \sqrt{(Q^2 + U^2)} \qquad 7.52$$

$$P = \frac{100.I_p}{I} \qquad 7.53$$

$$\theta = \tfrac{1}{2} \arctan \left({}^{U}/_{Q}\right) \qquad 7.54$$

In deriving the probable errors on these quantities, it is convenient to assume that the errors on I,Q,U are independent. This is true to a good approximation, since Q and U are derived primarily from different pairs of electronographs (Equations 4.21, 4.22), and the errors on I contain components due to other causes which do not affect Q and U. Hence the errors can be found from:

$$\Delta I_p^2 = \left(\frac{\partial I_p}{\partial Q}\right)^2 \Delta Q^2 + \left(\frac{\partial I_p}{\partial U}\right) \Delta U^2 \qquad 7.55$$

$$= (Q^2 \Delta Q^2 + U^2 \Delta U^2) / I_p^2 \qquad 7.56$$

and similarly:

$$\Delta P = P\sqrt{\left(\left[\Delta I_p / I_p\right]^2 + \left[\Delta I / I\right]^2\right)} \qquad 7.57$$

$$\Delta\theta = \frac{\sqrt{(Q.\Delta U)^2 + (U.\Delta Q)^2}}{(Q^2 + U^2)} \qquad 7.58$$

In some cases, when the errors are large, the linear approximations made here are not valid. In such cases,

however, the corresponding measurement is usually of little use and an accurate reflection of its error is not needed.

The complete dataset describing the polarisation measurements at the end of the data reduction procedure consists of a list of the quantities:

$X, Y, I, \Delta I, Q, \Delta Q, U, \Delta U, I_p, \Delta I_p, P, \Delta P, \theta, \Delta\theta$

where X and Y are the coordinates of the measurements in the photocathode frame of reference. Suitable transformations can then be made to convert these measurements into standard reference frames as suits the application.

7.5.9 Plotting the Polarisation Map

The data may be converted to a form for visual appreciation by plotting a line at each X,Y location of length proportional to the % polarisation at that point (or occasionally polarised intensity) and with the orientation corresponding to the direction of the electric vector. The resulting polarisation map may then be compared with a photograph of the object under study.

In plotting such maps, it is useful to be able to omit some measurements with large errors to give a visually appealing display. According to the nature of these errors, limits may be imposed on the measurements plotted, e.g.

Upper limit on I - Prevents stars being measured.

Lower limit on I - Prevents sky signal noise being plotted.

Lower limit on I_p - Supresses data with low polarised signal.

Upper limit on $\Delta\theta$ - Supresses data with large angular error.

Usually a combination of such limits is appropriate, as the errors depend in a complex way on the brightness, degree of polarisation and orientation of the polarisation in different areas of the map.

CHAPTER 8

SUMMARY OF THE DATA REDUCTION PROCEDURE

8.1 Introduction

Apart from the photocathode calibration stage, figure 8.1 shows an outline of the data reduction procedure, leading from the electronographs to the plotting of the final polarisation map.

Throughout the development of the reduction procedure, emphasis has been placed on automation, so that the user of the reduction programmes has to make the minimum of interaction with the computer. This results in less likelihood of human error, and since the programmes provide comprehensive printout on all decisions made it is easy to check for possible errors at a later stage.

Automation also brings the important advantage of speed - the complete reduction can be performed overnight if necessary - and this ability to rapidly repeat calculations which involve many stages of decision means that the effect of various changes to the technique can be quickly evaluated.

The data reduction is handled by five programmes which operate in sequence on the data (Figure 8.2). Results calculated at one stage which are relevant to later stages are passed on as part of the output dataset. The function of the programmes will be described briefly

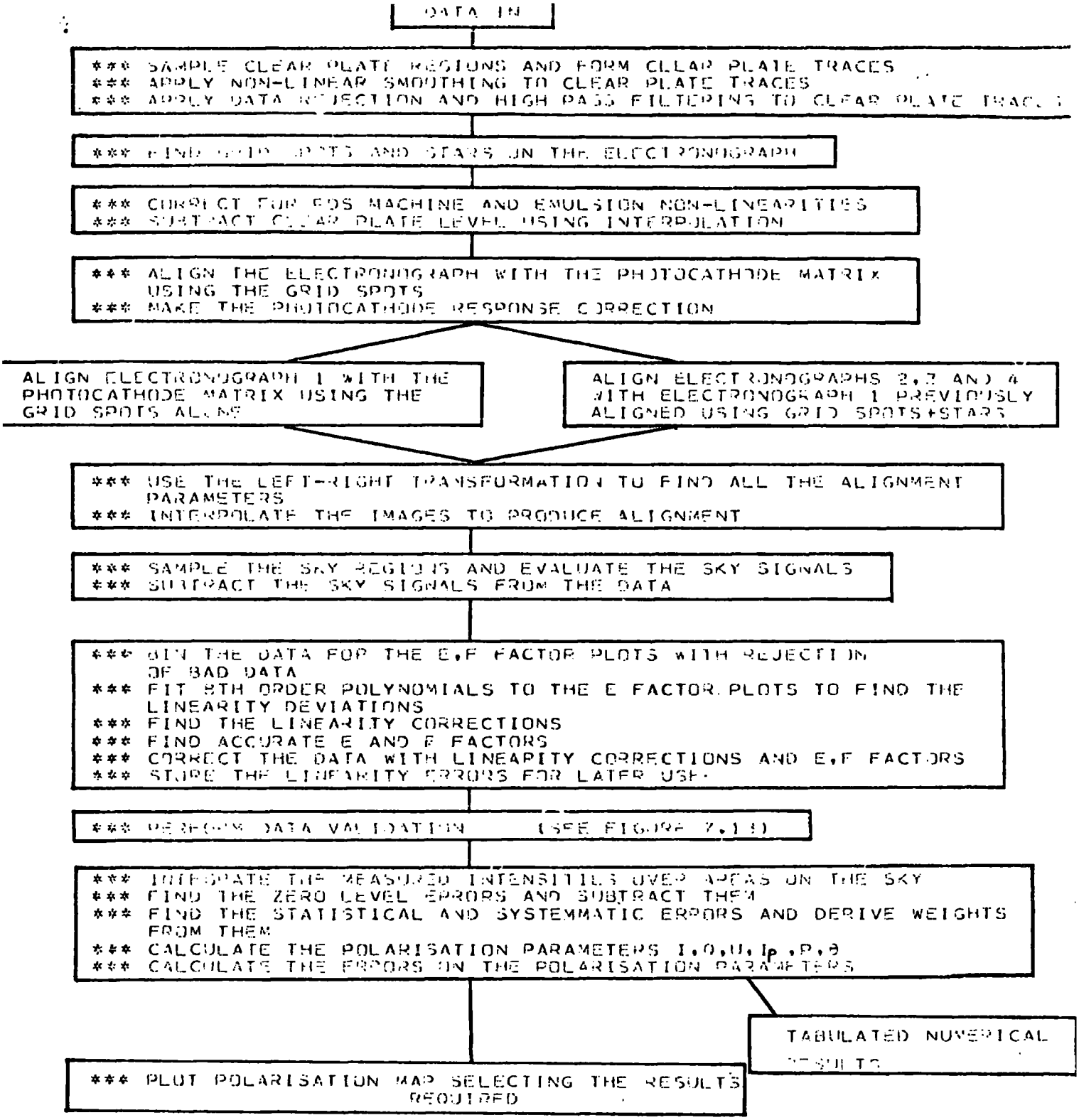


Fig.8.1 Summary of the data reduction procedure

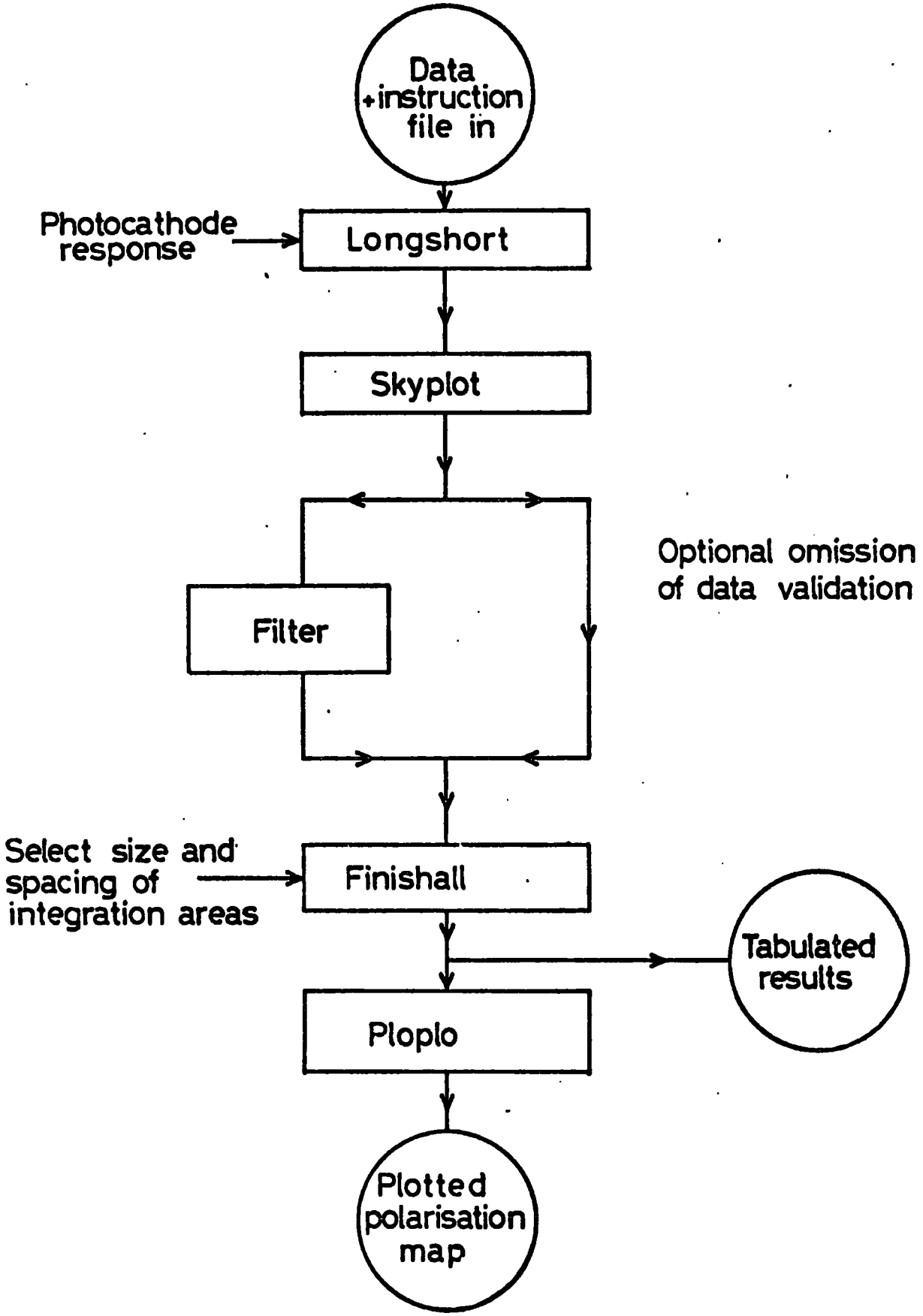


Fig.8-2 Sequence of data-reduction programmes

in order.

8.2 The Reduction Programmes

8.2.1 Longshort

The user specifies the approximate location of the star images in the raw data (e.g. Figure 3.3) and the regions in the image from which the sky signal is to be sampled (e.g. Figure 5.15). In addition, there are a number of control flags to be set indicating such things as the type of nuclear emulsion used for the data, any special alignment required (there is facility for aligning two polarisation maps for comparison purposes) and whether an object or photocathode calibration run is required.

The programme is provided with up to 4 data files, and it then handles the following steps in the data reduction.

i) Location of 20 grid spot images on each data file (5.2.4)

ii) Location, sampling and smoothing of clear plate signals (5.1)

iii) Alignment with the photocathode response (5.2.3)

iv) Subtraction of the clear plate level, and application of linearity corrections (5.1.6) and photocathode response correction (5.2.7)

v) Location of Stars

vi) Mutual alignment of the four electronographs (5.3)

vii) Sampling and subtraction of the sky background (5.4)

viii) Re-writing the data in compact form in the output file.

ix) Display of all results and decisions relevant to the running of the programme.

A full run of this programme uses ~200 seconds of CPU time.

8.2.2 Skyplot

The user supplies information relating to the number of electronographs being processed (fewer than 4 may be used to give somewhat less accurate results than with a full set). The scale factor F' is also supplied, derived from other data, in cases where it cannot be found due to missing electronographs. The user also specifies whether linearity corrections (6.3, 6.4) are required. The programme handles the following:

i) Logic to decide which plots (6.2.3) are possible, and which scale factors cannot be found.

ii) Binning of the data, and rejection of bad data (6.2.4)

iii) Fitting polynomials to determine the linearity deviations (6.3.2)

iv) Determination of accurate scale factors (6.4.1)

v) Calculation of linearity corrections (6.4.1)

vi) Correction of the data with linearity corrections and scale factors

vii) Display of all results obtained, including line-printer pictures of the corrected data.

A full run of this programme uses ~80 seconds of CPU time.

8.2.3 Filter

This programme handles the data validation process. Because of the nature of this process and the symmetry required in the data (7.3.4), this programme will only operate on a full set of 4 electronographs. It can, however, be omitted and the reduction procedure completed without data validation if required. No user input is required and the operation is described in chapter 7 and summarised in figure 7.13. A full run of this programme, using 5 iterations, takes ~150 seconds of CPU time.

8.2.4 Finishall

This programme handles the conversion of output from "Skyplot" or "Filter" to polarisation parameters, including the calculation of errors. The user supplies the dimension and spacing of the integration areas on a rectangular grid. The programme determines from the nature of the input whether data validation and linearity corrections have been applied. If they have not, the error calculations are omitted or modified, and a message

printed to this effect.

The time required by this programme depends on the size and number of the integration areas.

8.2.5 Ploplo

This programme handles the plotting of polarisation maps and is designed for interactive use on a computer graphics terminal. The user specifies a number of parameters such as the limits to be imposed on the data (7.5.9), the size of plot required and the region to be plotted.

There is also provision for plotting maps of polarised intensity instead of the more usual % polarisation and for including error bars on the maps. A polarisation scale, and the stars within the field used for alignment are also displayed automatically.

CHAPTER 9

ACCURACY AND FILM TYPE

9.1 Introduction

Having now described the data reduction procedure in full, we turn to a discussion of the sources of error apparent in the polarimeter system, with a view to possible increases in accuracy in future use. Since the discussion will deal with the overall accuracy of electronography, it is applicable to other applications apart from polarimetry, which simply serves as a means of investigating the errors in this chapter.

9.2 Random Noise

As has been mentioned previously, the evaluation of the amount of random noise on the polarimeter signal is seriously affected if we do not know the sensitivity of the film being used (7.4.2). In the case of films thought to be of good sensitivity (see discussion below - section 9.5.1), the amount of random noise present is of the order expected from photon noise alone (Figure 7.17), and this suggests (although it cannot be considered conclusive at this stage) that electronography may be capable of as high an accuracy in this respect as true photon-counting systems. (To establish this for sure, accurate calibration of the film will be necessary).

The overall errors from photon noise will thus be affected most by the number of photons "lost" in the system, i.e. by the quantum efficiency of the electronographic camera photocathode. Thus it is quite logical that highest accuracy will be obtained by using a camera with a high quantum efficiency.

9.3 Calibration Errors

Even in a true photon-counting system, the total error is not due to photon noise alone, but also to errors in the spatial calibration of the quantum efficiency of the detector. To some extent, inaccuracies in the spatial response calibration are unimportant in the Durham polarimeter, since to first order the effects are made to cancel by appropriate choice of the recorded images. However, whenever background subtraction is appreciable, spatial calibration becomes the major source of error once again. The accuracy achieved here depends essentially on the degree of reproducibility of this spatial response, i.e. on its stability, and on the accuracy of the technique used for calibration.

In this respect, electronography may be at a distinct advantage over some other systems which use detectors not designed specifically for astronomy, since questions of stability and rigidity of construction have been taken into account when developing the electronographic camera.

Practical experience of the calibration procedure (5.5) shows that errors of technique can account for quite large degrees of mis-calibration, and it is for this reason that considerable effort has been put into developing a suitable procedure. Checks on the accuracy of this calibration (5.4.5) show that when background effects (to be discussed later - section 9.5.2) are negligible, accuracies better than 0.5% are achieved. This compares well with the accuracies of other imaging detectors in this respect, and it would seem that electronography may be a potentially very accurate detector system.

9.4 Other Systematic Errors

Unfortunately, the above discussions do not paint a complete picture of electronography, which is usually regarded as inferior to all-electronic detectors in a number of respects, namely:

i) Relative difficulty of use and liability to faults.

ii) Lower overall accuracy.

iii) Inferior linearity to photon-counting systems.

While there is undoubtedly some truth in the first criticism, as the complexity of the reduction technique shows, the latter criticisms may not be due to inadequacies inherent in electronography itself, but to inadequacies

of the film used in the McMullan camera, as the following discussion suggests.

9.5 The Electronographic Film

9.5.1 Film Sensitivity

Tests have been carried out to determine the relative sensitivities of different batches of film by making equal exposures with a constant light source and comparing the photographic densities after development. Due to the lack of measuring equipment at most observing sites, this comparison has been made by eye in most cases, but the results are sufficiently obvious, even to the inexpert eye, that it is possible to determine "good" and "bad" batches by this technique.

Variations in sensitivity, as estimated by eye, of factors of 2 seem commonplace, with factors up to 5 being encountered at times. Recently, laboratory tests have been made at the Royal Greenwich Observatory on samples of batches actually used (these were selected by eye as "good"), and the results are summarised in table 3:

Emulsion Batch	Relative Sensitivity
PL 595	1.3
PL 698	1.0
PL 703	2.1
PL 707	1.8
PL 727 '10μ'	2.3
PL 727 '15μ'	2.0

Table 3

The occurrence of such variations raises some doubts about the effect they may have on accuracy, and it is unfortunate, also, that until recently most users of the McMullan camera have been unaware of these variations. While the loss of sensitivity is undoubtedly a feature of the film manufacturing process, it seems unlikely that the most obvious reason - changing emulsion thickness - can account for it, for several reasons. Firstly, the sensitivity is fairly constant for a given batch (the batch refers to the emulsion, not the coating of the melinex backing, which may be performed later), and secondly, if thickness variations were responsible it would imply that electrons could penetrate the emulsion completely and this would cause other serious errors depending on the exact thickness of the emulsion at each point. (If the electrons do not penetrate the emulsion layer - the density is independent of the emulsion thickness). Finally, coating the film with a 15μm emulsion layer, instead of the usual 10μm layer, gives no dramatic improvement in sensitivity (Table 3, batch PL 727) which cannot be explained as the variation found within a single emulsion batch. The probable source of variation lies in the constitution of the emulsion itself, and practical tests have shown that a "good" batch can usually be recognised, as the unprocessed film shows a higher opaqueness to transmitted light.

In all cases, it is desirable to have a sensitive batch of film on which to record faint images, since information recorded at low photographic densities is more difficult to measure accurately, thus all film used with the Durham polarimeter is subjected to a sensitivity test before being used. However, since the absolute sensitivity of the film is not usually known, the calculation of theoretical linearity corrections (4.2.2) may be in considerable error, possibly explaining criticism (iii) (9.3).

9.5.2 Errors Due to Film Quality

There are other features of nuclear emulsion, well known to electronographers, which give rise to errors. The most renowned of these is contamination by dust and scratches, and the treatment of this problem has been dealt with in chapter 7. In addition, the visual appearance of the film is rarely uniform, showing regions of varying emulsion thickness. The effect of this on the developed density is not known, but changes in thickness of the developed film of only 1% could change the background density by ~ 0.5 PDS units, resulting in significant error.

Coupled with these errors due to the usually poor quality of film, there are others which can arise during film processing, e.g. drying stains and marks on the melinex backing, etc.

9.5.3 Features Required of Electronographic Film

Undoubtedly, to achieve the highest possible accuracy with electronography, film of a much higher quality must be used, and handling and film processing must be carried out with extreme care. The main problem to be overcome is that of availability, since film manufacturers are understandably reluctant to invest in equipment for high quality film production for a use which provides so small a financial return.

In the current absence of an ideal alternative, some investigations have been made into ways of best using the available film. Since practical results suggest that there is an inherent background error of $\sim$0.5 PDS units on L4 emulsion, the possibility arises of using a film of higher sensitivity, such that signals are recorded in a density range where the background error is negligible.

9.6 Comparison of L4 and G5 Emulsions

9.6.1 Introduction

There is a film of higher sensitivity available using G5 nuclear emulsion, however L4 is usually preferred on the grounds of better cleanliness, higher resolution (G5 has larger grains) and wider dynamic range. There are often instances, however, particularly with faint objects, where resolution and dynamic range are not needed, while an ability to record faint objects with

reasonable densities is important in order that the signal should be well above the inherent background errors.

Since G5 has a sensitivity nominally 4 x that of L4, the background errors could potentially be reduced by a factor of 4 for equal exposure times. The possible limiting factor is that if the errors are due to emulsion variations, they may scale in the same way as the recorded density, while if they are true "background" errors, the full advantage may result. To test this, a comparison has been made by producing two polarisation maps of the same object using the two film types.

9.6.2 Observations for the Film Test

The object chosen for this test, NGC 6726/7, is a reflection nebula associated with the stars TY CrA and HD176386 in Corona Australis. It has a wide range of surface brightness, shows high polarisations and has a polarisation pattern which is smooth in some regions and rapidly changing in others, as such it provides a good test for the characteristics of the two film types over a wide range of measurement.

Previous polarimetric observations have been made in this region by Serkowski (1969) using a photo-electric instrument, and the reader is referred to the literature for a further discussion of this object. In this context it will be discussed only as a test object for investigating

the two film types.

The observations were made during the night of 3/4th July, 1978 using the 1 metre telescope of the South African Astronomical Observatory at Sutherland, S.A. The exposures were 15 minutes per electronograph for both film types, using the V waveband of the UBV system. Observing conditions on this night were excellent: the sky dark (no moon), the object well above the horizon and no indications of cloud. To ensure similar conditions for both film types, the data for comparison purposes were collected during successive 1 hour periods using exactly the same polarimeter position on the object. The resulting polarisation maps have been accurately aligned with each other using the stars within the field so that the placing of the integration areas on the sky is identical in both cases. The regions used for sky subtraction are similarly identical for both film types, so that the results are directly comparable in all respects.

9.6.3 The Polarisation Maps of NGC 6726/7

Figure 9.2 shows the polarisation maps produced using L4 emulsion. The integration areas are 5 x 5 pixels, corresponding to an area on the sky of 6.9 arc seconds square. The interlacing of the grids IN and OUT measurements give the characteristic strips of measurements with the gaps corresponding to areas of image rejected due to overlapping and other effects.

All the measured values are shown on this map (except for the 6th grid aperture, which was used for sky signal determination), so that regions of noise are not supressed. Figure 9.1 shows the same map without data validation so that the effect of this technique can be seen.

The map shows an unmistakable pattern of polarisation vectors with polarisations ranging from zero to ~30%. The pattern is consistent with polarisation produced by light scattering within the nebula, illumination being provided by both the stars TY CrA and HD176386. This double illumination induces neutral points in the polarisation pattern (N and N′ in figure 9.2), where the two stars subtend an angle of 90^{o} and the polarisations produced by scattering from the two sources are orthogonal and thus cancel. Also seen are regions where the measurements are rather uncertain due to the faintness of the object.

Only the grids IN half of the object has been measured using G5 emulsion for comparison, and the resulting map is shown in figure 9.3 where it has been interlaced with the corresponding results using IA emulsion such that identical regions of the nebula are presented side by side for comparison. The integration areas are again 5 x 5 pixels in both cases.

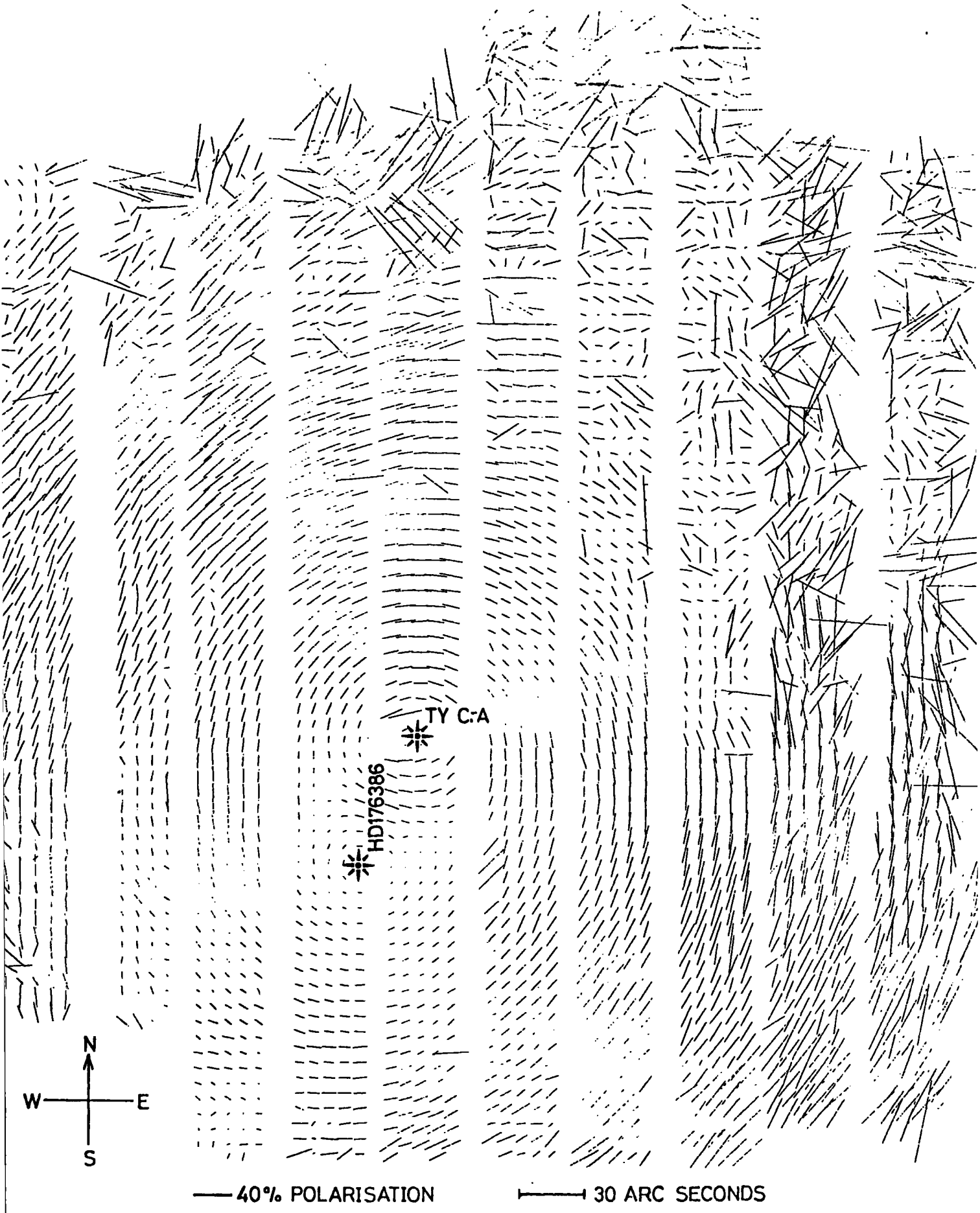


Fig.9.1 The map of NGC 6726/7 before data validation

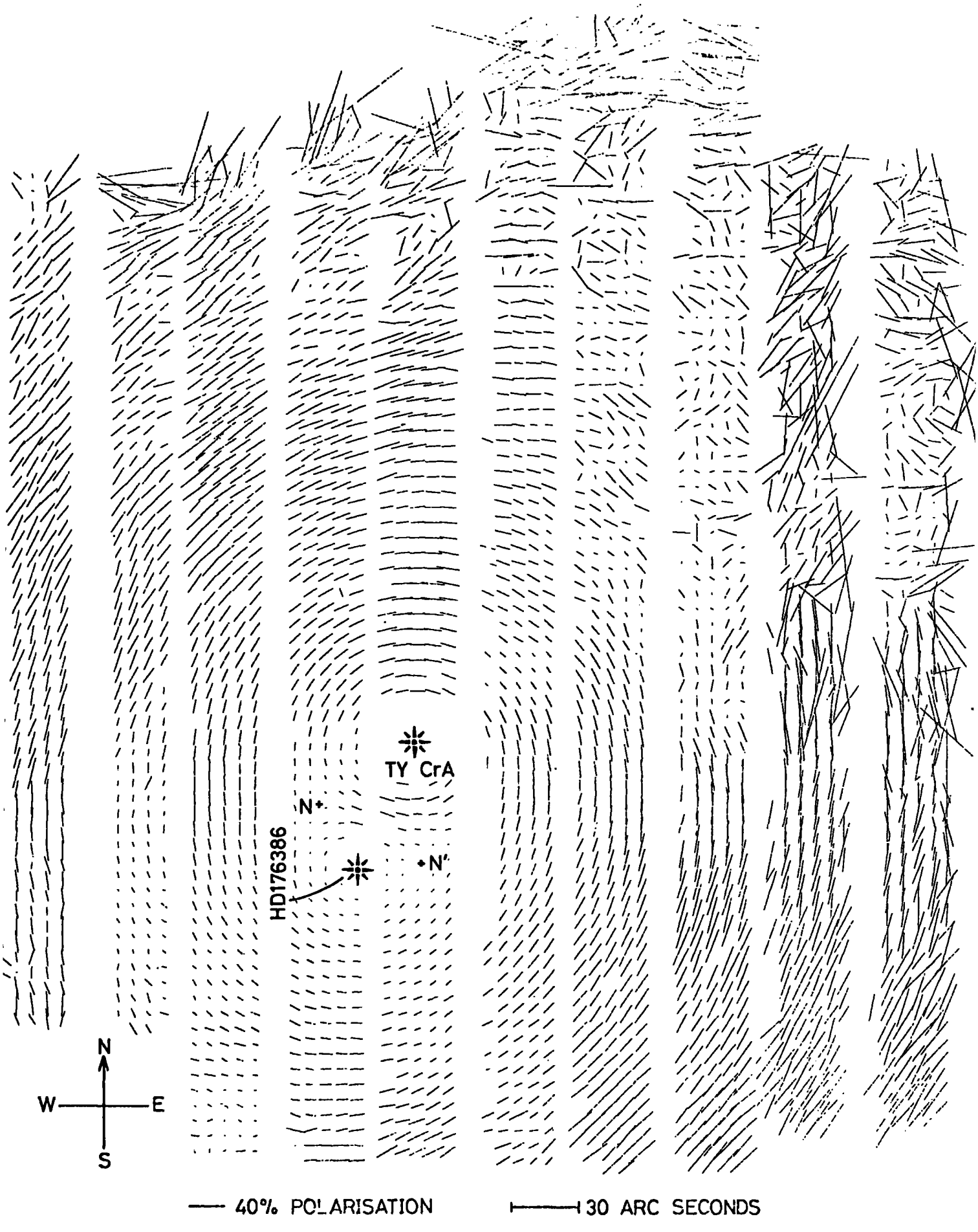
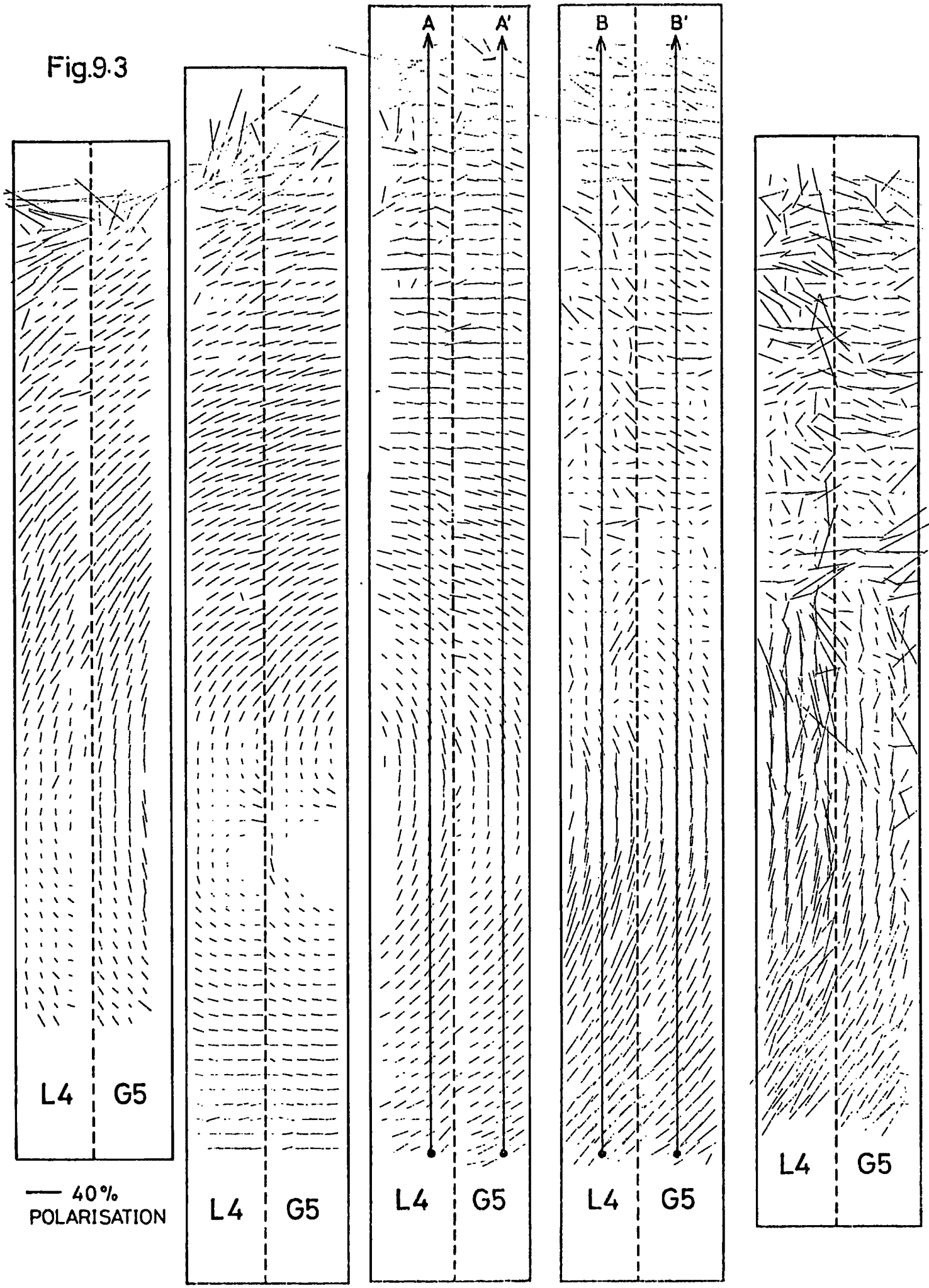


Fig.9.2 The polarisation map of NGC 6726/7 produced using L4 emulsion

Fig.9.3



9.6.4 Results of the Film Test

A comparison of the two maps in figure 9.3 shows that over most of the field, where the object is relatively bright, there is no significant difference between the L4 and G5 maps, except where the greater dynamic range of L4 shows as a smaller region of over-exposure around the star HD176386. This is consistent with the expected domination of statistical errors at high recorded densities, since these are primarily due to photon noise (9.2) and are thus unaffected by the emulsion type.

There are, however, some regions where a significant difference is seen, notably at the top of the 3rd and 4th grid apertures and at the bottom of the 5th. These regions correspond to the fainter parts of the object, and it is in such regions that systematic errors should be important and possible differences be seen. In all these regions the G5 map shows a smoother and more noise-free appearance, consistent with the expected fact that recording the signal at higher density on G5 emulsion makes the systematic errors less significant.

To investigate this difference more fully, figures 9.4 and 9.5 show traces of total and polarised intensity through a faint region of the object (along the lines A, A′ in figure 9.3) together with the calculated errors. The vertical displacement of the L4 and G5 curves corresponds to the relative sensitivity factor of 4, and the error bars, on logarithmic plots such as these, are proportional,

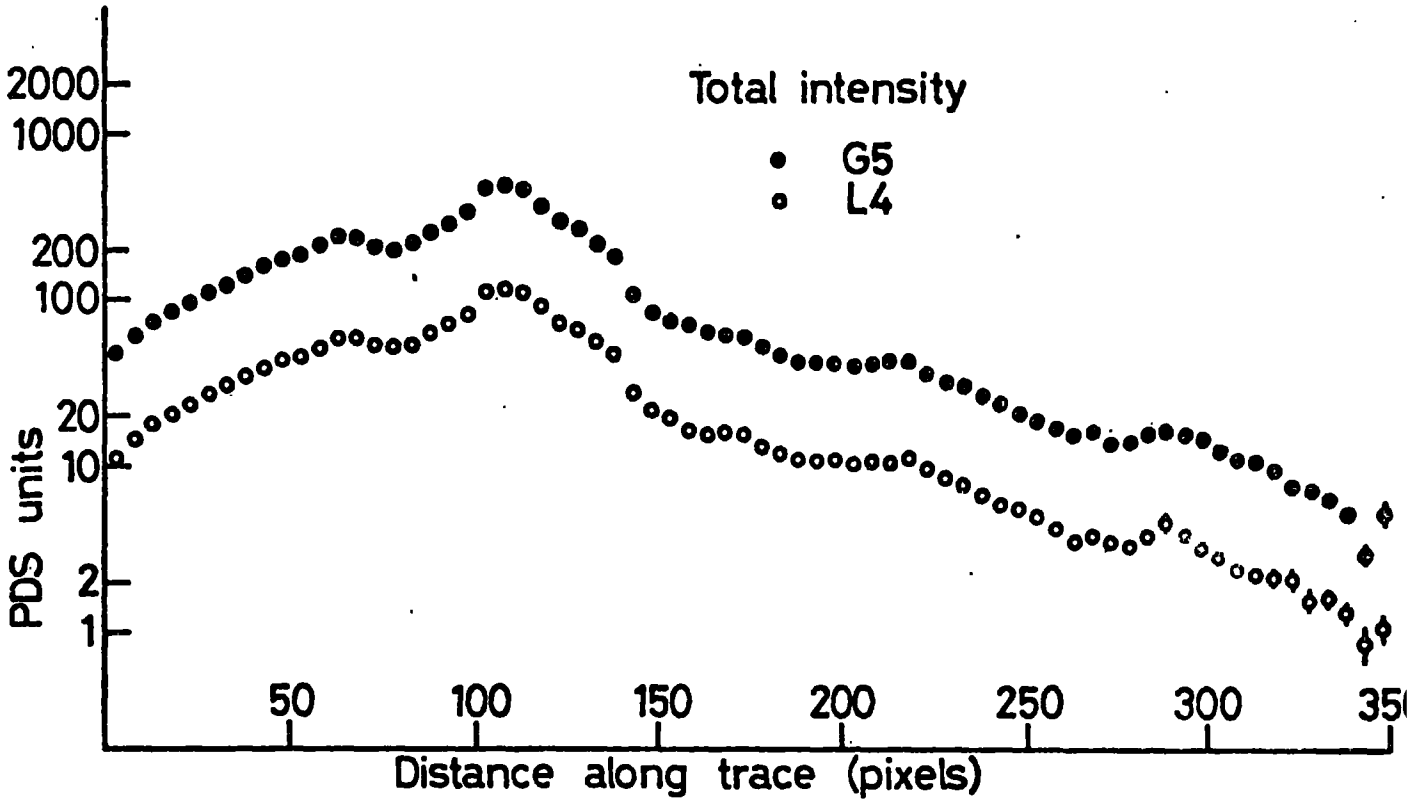


Fig.9.4 Total intensity along trace A (fig.9.3)

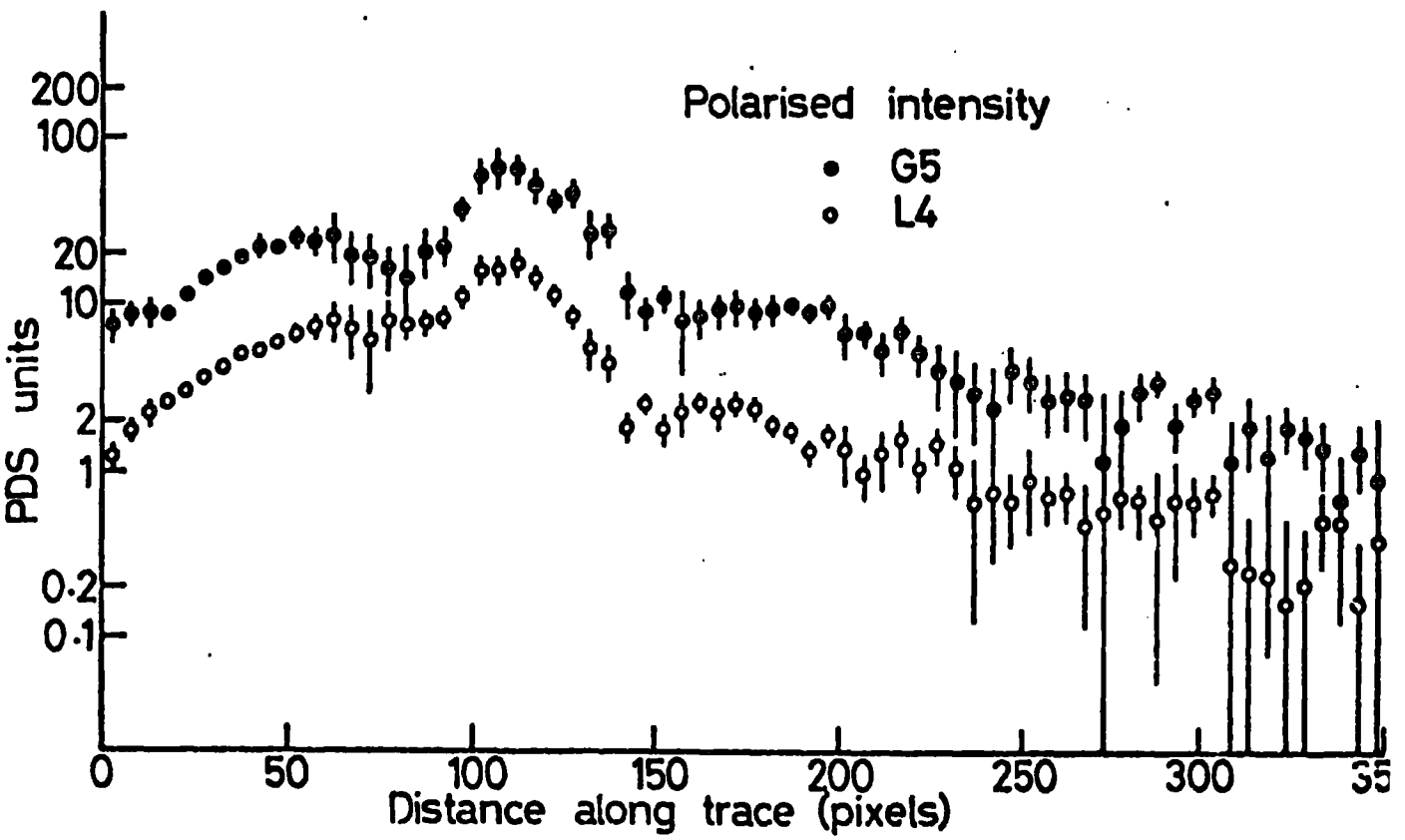


Fig.9.5 Polarised intensity along trace A (fig.9.3)

in absolute length, to the fractional error on the quantity being measured.

The quantity relevant to the measurement of polarisation, polarised intensity, shows little difference between the L4 and G5 results at high signal levels (i.e. high total intensity), but the calculated fractional error on the G5 results are significantly less at low signal levels. If this increasing difference at low signal levels is indeed due to systematic errors, a greater effect should be seen when using large integration areas, such as when observing very faint objects, since the errors will then be due entirely to systematic effects. Figure 9.6 shows traces of polarised intensity through a very faint region of the object (traces B, B′ in figure 9.3) using integration areas of 10 x 10 pixels. The effect is indeed more pronounced, with the fractional errors on G5 emulsion being at least halved with respect to L4.

9.7 Overall Accuracy

Since increasing the area of integration can only increase the accuracy of measurement up to the point at which systematic errors become important, it is necessary to know the size of these systematic errors so that the ultimate accuracy of the system can be estimated.

Figures 9.7 and 9.8 show scatter plots of the error on the polarised intensity against the total intensity for L4 and G5 emulsions, derived from the data in figure

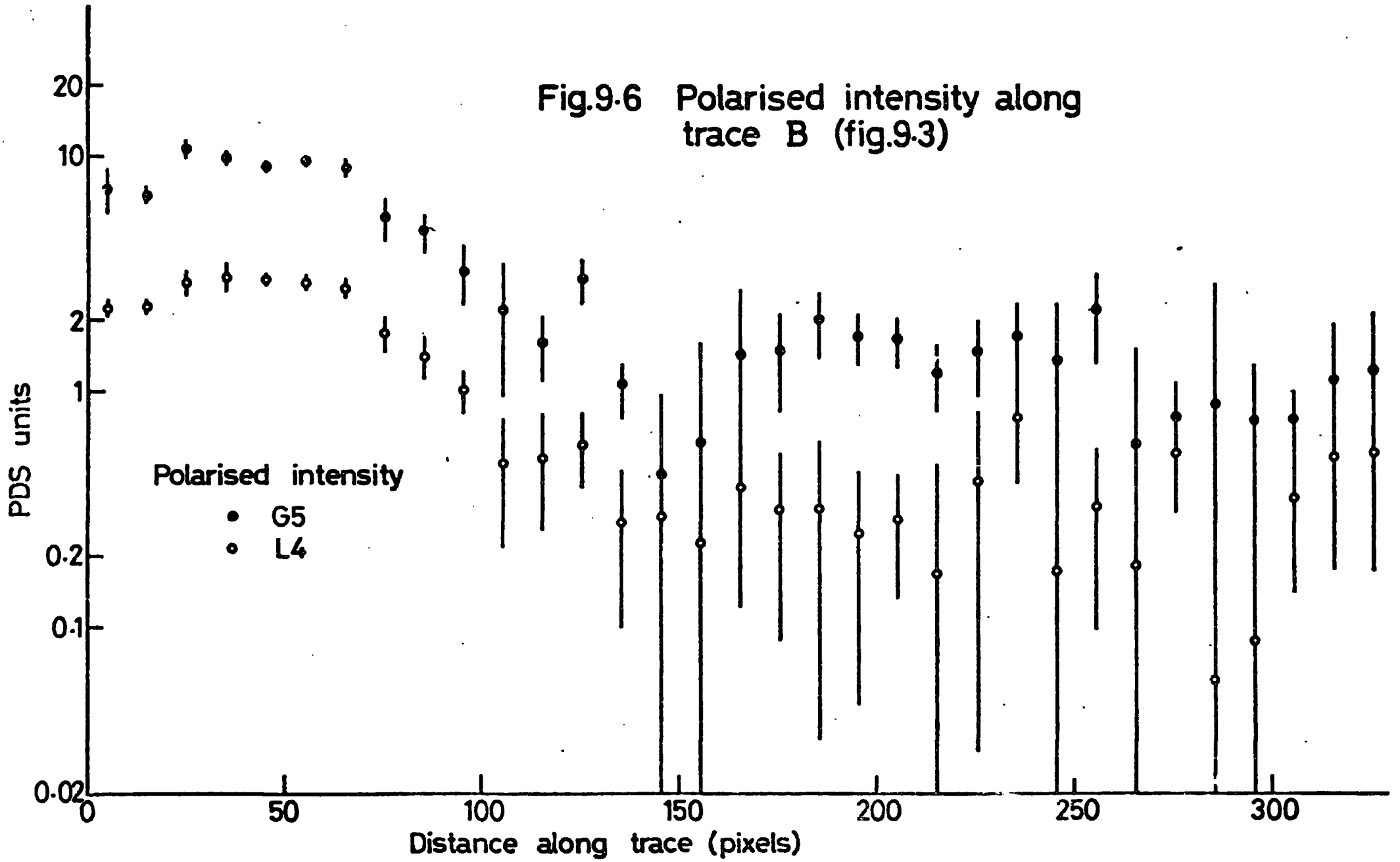


Fig.9·6 Polarised intensity along trace B (fig.9·3)
PDS units
20
10
2
1
0·2
0·1
0·02
Polarised intensity
G5
L4
0
50
100
150
200
250
300
Distance along trace (pixels)

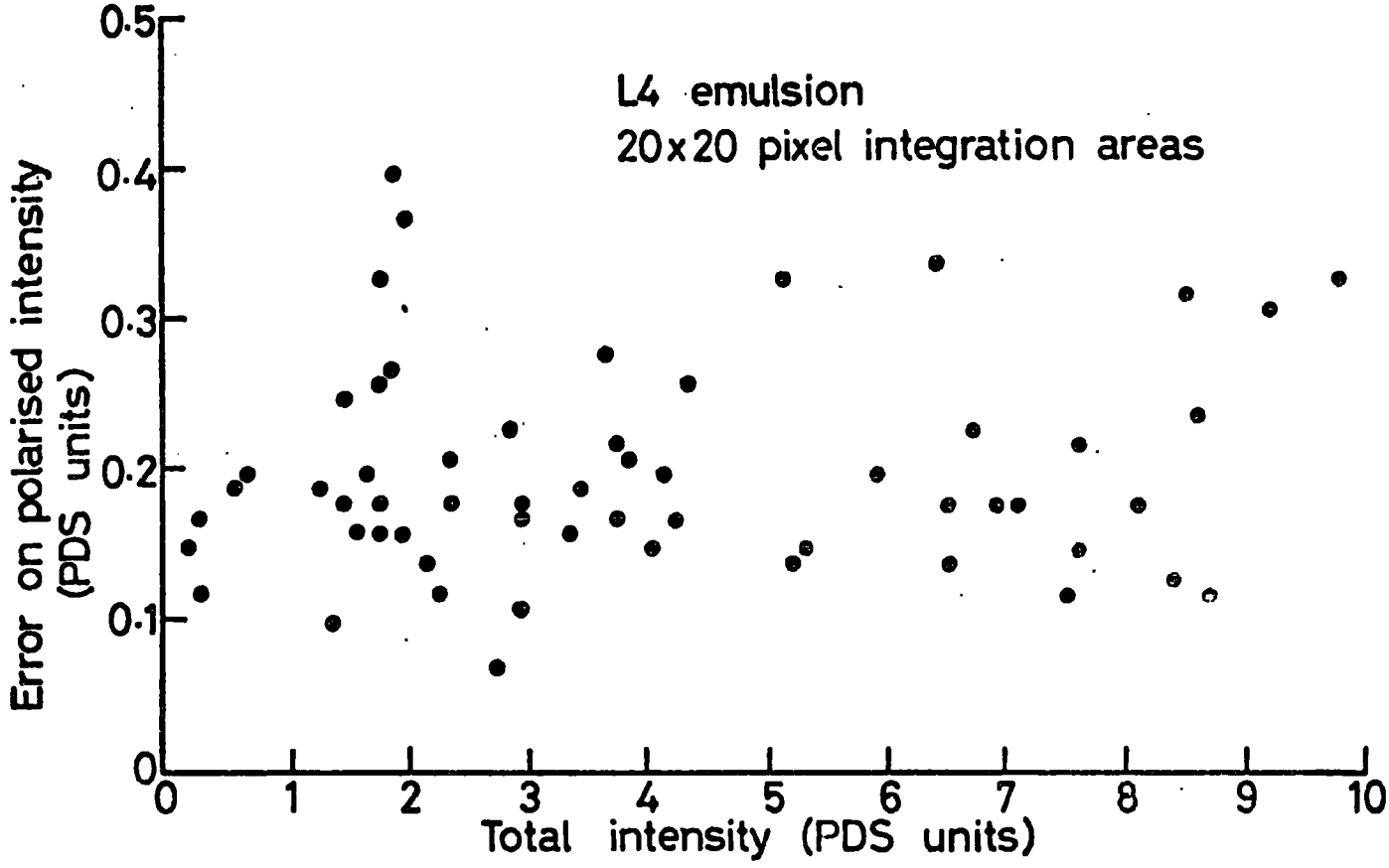


Fig.9·7 Error plot for L4 emulsion

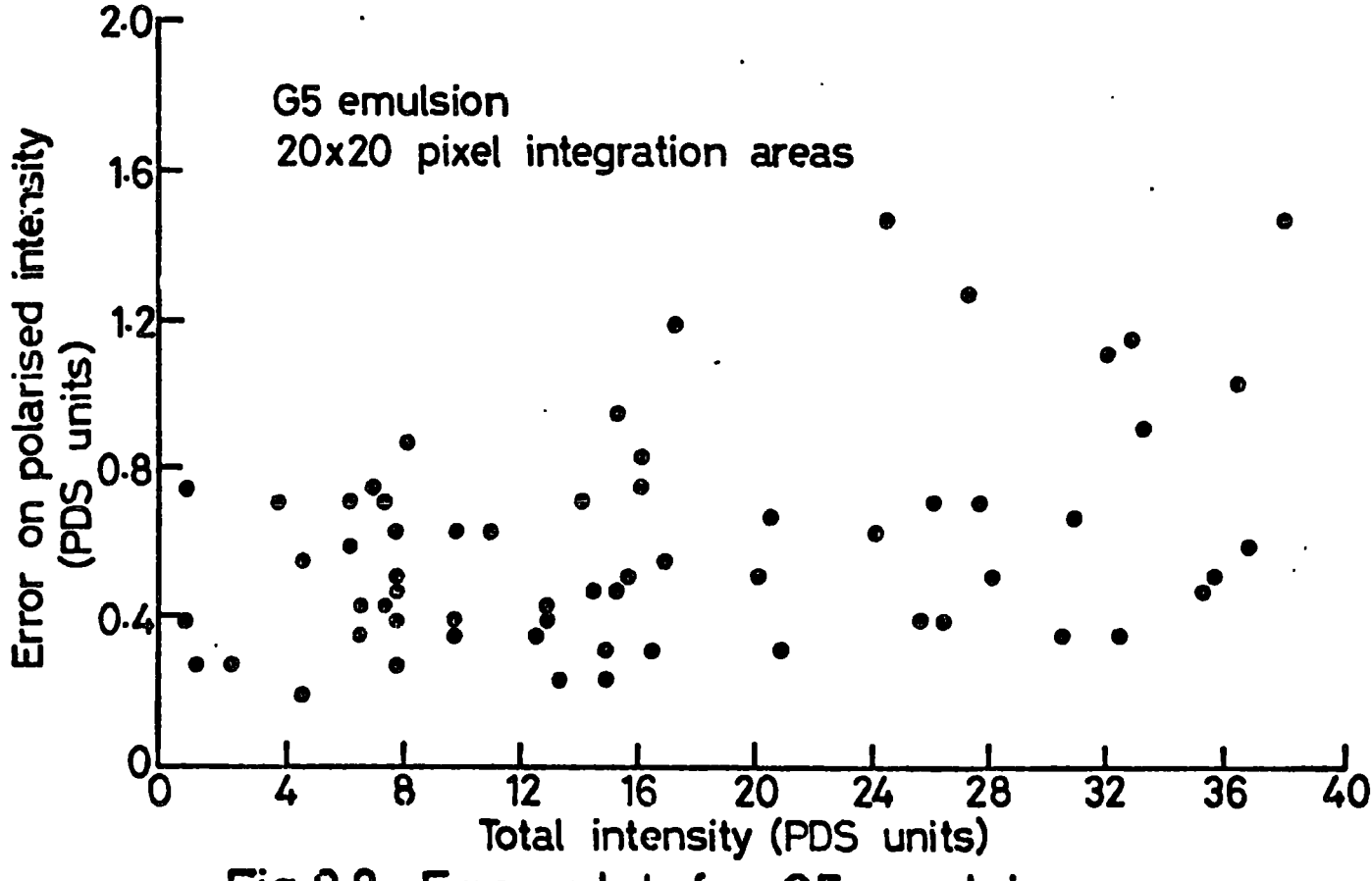


Fig.9·8 Error plot for G5 emulsion

9.3, but using integration areas of 20 x 20 pixels to minimise the contribution from statistical error. By extrapolating these plots to zero intensity, the intrinsic background errors on the measurement of polarised intensity can be found for the two emulsion types. The values found are 0.17 and 0.4 PDS units for L4 and G5 respectively (the expected random contribution is negligible in comparison).

Since the background errors on G5 emulsion exceed those on L4, the full advantage of a factor of 4 is not obtained when using G5 (9.6.1), however, some advantage is still obtained, amounting to a factor of ~ 1.7 in ultimate accuracy when using G5 in the case of the test described here.

It must be appreciated that the factor of 1.7 is obtained only when using large integration areas and faint objects, since the presence of statistical errors serves to make the two emulsions more nearly equal in accuracy. We can combine the intrinsic background errors derived here with the theoretical estimate of statistical errors (Equation 7.4) to give an approximate expression for the expected error on the polarised intensity ΔI_p when the total intensity is I, and the sky background signal subtracted is S, using an integration area of mxn pixels:

$$\Delta I_p = \sqrt{\frac{0.08(I+S) + 0.083}{(m+1)(n+1)} + 0.029} \qquad 9.1$$

for L4 emulsion, and:

$$\Delta I_p = \sqrt{\frac{0.32(I+S) + 0.083}{(m+1)(n+1)} + 0.16} \qquad 9.2$$

for G5 emulsion.

(Where equations 4.21 and 4.22 have been used in estimating the error.)

We can also relate the intrinsic background errors derived here to the ultimate sensitivity of the technique to measurements of polarised intensity. Absolute calibration of this test field is difficult, but if the sky brightness is used as a reference, and assumed to have a value of 21.5 V magnitudes/square arc second, and if the polarised intensity is to have a value twice the background errors in order to be detected, the minimum detectable polarised intensities are:

25.8 V magnitudes/sq. arc sec. on L4 emulsion

26.4 V magnitudes/sq. arc sec. on G5 emulsion

for the case of the 15 minutes per electronograph exposures used here. Accuracy would be expected to increase linearly with exposure time, since systematic errors are dominant at low surface brightness.

At higher surface brightnesses, it is difficult to describe the accuracy concisely, since it depends on many factors such as the integration area, % polarisation, orientation, etc. However, some idea of overall accuracy can be gained from the knowledge that for 90% of the measurements in figure 9.2, the fractional error on the percentage polarisation (i.e. $\Delta P/P$) is less than 0.3, and the error on orientation less than 8^{o}. Typical errors for this map are somewhat less than this, with $\Delta P/P \sim 0.1$ and $\Delta\theta \sim 4^{o}$.

Axon (1977) has shown, using measurements of standard polarised stars, that the optical system of the polarimeter introduces negligible instrumental error when polarisations of $\sim$1% are measured. Since the errors calculated within the data reduction system are rarely negligible at this level, it is thought that they represent the true errors in measurement, the instrumental effects being insignificant by comparison.

SUMMARY AND CONCLUSIONS FOR PART 1

The technique described in the preceeding chapters has been developed for the production of polarisation maps from images produced by the Durham polarimeter and recorded using a McMullan 4 cm. electronographic camera. The computer techniques used have been designed to facilitate data reduction with the minimum of user intervention in the routine measurement of polarisation in extended astronomical objects, and give reliable and consistent results down to low brightness levels.

Significant improvements in accuracy over previous methods have been achieved by careful checking of the data at all stages of processing to eliminate the effects of data corruption. Sky signal subtraction, image normalisation and spatial calibration of the detector response (flat-fielding) are particularly sensitive to image corruption and improvements have been made in all these areas. The accuracy of zero level subtraction has also been greatly improved by the development of a technique to separate real variations in clear plate density from image defects and drift in the PDS micro-densitometer. Accurate and reliable image alignment has been made automatic by the use of a set of fiducial marks which allow good image registration to be achieved even if only a single star is available in the polarimeter field. A large degree of automation allows the complex data reduction to be accomplished quickly, and comprehensive printout on all stages of processing allows rapid diagnosis

of data errors and related problems.

A new technique, based on fitting 3-dimensional functions to the data, has been developed to check the internal consistency of the polarisation data, both as regards smoothness within each image and mutual consistency between the 8 images used. Use of this method to eliminate image blemishes gives improved estimates of the polarisation parameters, and a 2-dimensional version of the technique, developed initially for testing, gives good results when applied to single images. Making use of a measure of redundancy in the polarimeter data with the fitted 3-dimensional function (effectively comparing several images of the same object) enables estimates to be made of the error on each polarisation measurement from a variety of causes. In addition, the sources of error in electronographic recording can be investigated in some detail.

The errors resulting from image defects depend greatly on the quality of the data but, by using the above technique, they can usually be made acceptably small for most data. Spatial calibration of the detector response can probably be carried out to an accuracy better than $\pm$ 0.5%, but irreproducibility of the linearity of response between separate images may amount to 1 or 2%. This latter error may arise as a result of variations in film sensitivity which mean that the theoretical linearity corrections applied are not always appropriate

to the film actually used. At high density levels, polarisation measurement accuracy is limited primarily by statistical errors and the level of these is consistent (to within the accuracy of the film sensitivity calibration) with the expected level of photon noise, allowing for the quantum efficiency of the camera. There may be some contribution to statistical error at low density levels resulting from the line-by-line subtraction of the clear plate level - this error is acceptable, however, in view of the much larger errors which result if the zero level is not monitored in this way.

At low signal levels, errors are dominated by systematic effects with large spatial extent on the electronographs. These systematic effects may sometimes contain a significant contribution from clear plate subtraction errors occasioned by drift in the zero level of the PDS microdensitometer, but under most circumstances are attributable to non reproducible density resulting from non-uniformity of the film used. These effects are typically $\pm$ 1 PDS unit with L4 emulsion (1 PDS unit = a density increment of 0.005) and appear to be mainly additive (not multiplicative) in effect, and may thus result from the variations in emulsion thickness often apparent in the unprocessed film.

Electronographic recording appears able to achieve high accuracy as regards spatial calibration of the detector and low statistical noise levels at high recorded densities, but is considerably less accurate

as regards reproducible linearity and the effects of systematic background errors. These latter effects may be attributed to variations found in the film used and improvements in this area may lead to significant improvements in accuracy, particularly at low brightness levels. The areas of concern (in decreasing order of importance) are non-uniformity of individual film samples, inconstancy of film sensitivity between samples and the number of film defects. Where improvements cannot be made to film quality, use of the more sensitive G5 emulsion gives a significant improvement in accuracy over L4 (for equal exposure times) at low brightness levels. At high brightness levels, however, there is no significant difference between the 2 emulsions - consistent with the domination of photon noise - and the higher storage capacity of L4 is then to be preferred. The accuracy of measurement should increase linearly with exposure time at low brightness, but as the square root of exposure time for brighter objects.

PART 2

STUDIES OF REFLECTION NEBULAE

CHAPTER 10

OBSERVATIONS OF NGC1999 AND V380 ORIONIS

10.1 Introduction

The nebula NGC1999 is associated with the star V380 Orionis, which lies to the SE of the Orion Nebula in a region of dark obscuration and faint nebulosity, rich with T Tauri stars, Herbig-Haro objects and signs of young star formation within the Orion dark cloud.

NGC1999 is notable because of its high surface brightness relative to the illuminating star and also because of a striking triangular patch of dark obscuration seen silhouetted against the bright inner regions of the nebula, immediately W of V380. Due to this obscuration, the inner nebulosity takes on an arc-like appearance in short exposure photographs (Figure 10.1).

10.2 Previous Work on NGC1999 and V380 Orionis.

10.2.1 V380 Orionis

V380 Orionis is the illuminating star of the nebula NGC1999 and has received far more study than the surrounding nebulosity on account of its probable pre-main-sequence status.

Early observations by Jonckheere (1917) suggested that V380 might be double, or possibly triple, with companions of magnitudes 13.0 and 15 (V380 was reported

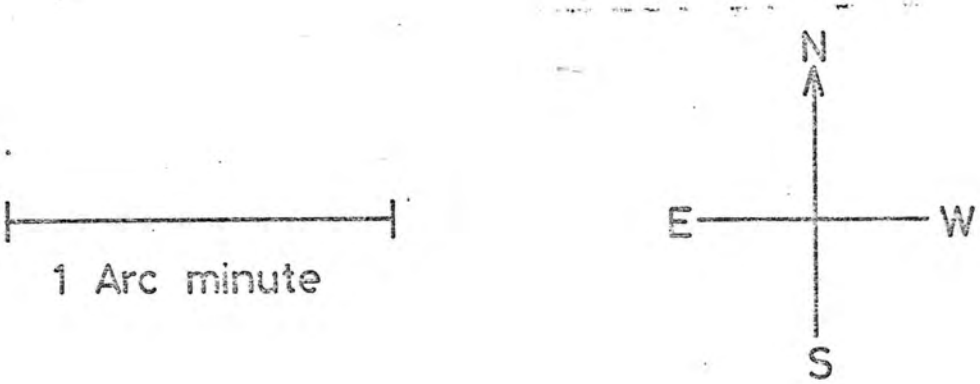


Fig.10.1 A mosaic picture of NGC1999 produced from polarimeter electronographs

as magnitude 8.9). No further observations have supported this claim, however, and V380 seems to have become accepted as having only one component. No evidence for companions has been found in the present polarimetric study, and it seems that the arcuate shape of the bright central regions of the nebula may have been a confusing factor.

V380 was selected by Herbig (1960), as one of a class of Ae and Be stars associated with nebulosity in dark clouds, which he considered to be candidates for being young massive stars still associated with the dust and gas from which they formed. A review of the properties of these Herbig emission stars, and features of other young stellar objects is given by Strom, Strom and Grasdalen (1975).

Spectra of V380 were discussed by Herbig (1960), and show numerous emission lines, those of H, CaII, FeII and TiII being prominent. On the strength of the Balmer wings in the underlying absorption spectrum, Herbig identified the spectral type as between B8 and A2. A tentative identification of absorption in the k line of CaII might indicate a spectral type of A0 to A2.

The emission line spectrum of V380 is complex compared to other members of the Herbig emission star class, and qualifies it marginally for membership of the class of T Tauri stars, although with a higher

degree of excitation corresponding to its early spectral type. P Cygni emission line profiles, seen in many Herbig emission stars, have not been observed in V380, although the lines are diffuse, with half widths of 66km s^{-1} for FeII, 95km s^{-1} for Hβ,Hγ and 125km s^{-1} for K of CaII (Herbig 1960). Garrison and Anderson (1977) find a half width of 154 km s^{-1} for Hα, with no indication of the asymmetry or double-peaked structure found in many other Herbig emission stars. The mean radial velocity derived from the emission lines by Herbig (1960), of +22km s^{-1} agrees well with the peak in the 21cm radio velocity measurements of the neutral HI cloud in this region of +24km s^{-1} (Menon 1958).

No evidence for spectral variation was seen by Herbig, although the star is thought to vary slightly in brightness. Measurements of its magnitude are confused, due to the variability and the unknown contribution from the surrounding bright nebulosity, but Herbig quotes a range from V = 9.7 to 10.1. Photometry of this star has also been conducted by Mendoza (1966, 1968), Racine (1968) and Kuhi (1974) and a discussion of the results will follow in section 10.2.3.

Brück (1974), in a photographic study of V380 and the surrounding nebulosity, derives colours of the nebulosity and attempts to resolve the confusion over the star's magnitude and colour. She shows that photoelectric measurements which include the bright inner

nebulosity should be corrected, and derives new B and V magnitudes based on fitting stellar profiles to the inner brightness profile of the nebula.

Observations by Mendoza (1968), Gillett and Stein (1971) and Allen (1973) indicate an infra red excess typical of the Herbig emission stars. Attempts to attribute this to free-free and free-bound emission from a circumstellar shell, and to thermal emission from circumstellar dust show that no distinction can be drawn on the evidence of the IR observations alone, although observations by Cohen (1973) which show an IR peak at ~10μm, possibly attributable to silicate grains, provides support for the presence of circumstellar dust.

In a polarimetric study of the Herbig emission stars, Garrison and Anderson (1978a) found evidence for intrinsic polarisation in V380. By a graphical method, which seeks to make the position angle of the intrinsic polarisation equal at all wavelengths measured, they determined the interstellar polarisation in the direction of V380 to be 2.0% at position angle 140° at a wavelength of 0.55μm. Their results, corrected for interstellar polarisation, are summarised below:

Waveband	U	B	V	R	Hα
% polarisation	1.06	2.18	2.31	1.77	1.42
Position angle °	72	63	65	68	74

Table 4

The presence of intrinsic polarisation suggests the presence of an asymmetric scattering medium around the star, possibly in the form of a circumstellar disc, while the lower polarisation in Hα suggests that the polarisation may be produced in the Hα emitting region, although the observations did not find this feature in all the Herbig emission stars. Comparison with previous work by Zellner (1970) and Breger (1974) shows evidence for variability in the intrinsic polarisation.

An extensive spectrophotometry survey of the Herbig emission stars was also undertaken by Garrison and Anderson (1978b) in the wavelength range 0.34μm. to 0.83μm. They derived an Orion-type reddening law correction, based on comparison of the observed spectrum of V380 with results from a theoretical model atmosphere for the star, using a temperature of 12,400K, and observed a smaller than expected Balmer decrement, in common with most of the Herbig emission stars observed (0.30 ± 0.05 compared with the expected 1.35 magnitudes). They attempted to explain the Balmer excess with free-bound emission from an optically thin circumstellar shell, and showed that this shell could also explain an observed excess at wavelengths above $\sim$0.58μm, if the temperature in the shell were $\sim 10^4$ K. This model, however, does not explain the excess beyond $\sim$1.5μm and thermal emission by dust with T$\sim 10^3$ K. is suggested to explain the IR. excess.

10.2.2 NGC 1999

The nebula NGC 1999 appears on long exposure photographs as an approximately circular object with a diameter of ~1.5 arc minutes and with the star V380 Orionis at its centre. On short exposures the dense obscuration gives the inner regions an arctuate appearance (Figure 10.1).

Herbig's spectra, which included several regions of the nebula, showed that the inner regions had a line spectrum identical to that of V380, while the outer regions showed some weakening of the lines of CaII, which might be attributed to absorption within the nebula. These observations lead to the general conclusion that the nebula shines by reflected light alone, intrinsic emission being absent.

The photographic study by Brück (1974) fitted the nebular data to an absolute scale of brightness using nearby photo-electric standard stars, and found that the nebular brightness and size satisfied the Hubble relationship :

$$m_{pg} + 4.9 \log_{10} a'' = 19.74 \qquad 10.1$$

relating the radius of the 23.4 magnitude per square arc second contour (a" = 47") to the photographic magnitude of the illuminating star. This relation is known to have bearing on the albedo of the scattering centres within the nebula, but a detailed interpretation in specific cases is difficult.

Photographic colour maps have also been produced by Brück (1974) with a resolution of 0.5 magnitudes in colour index. These results show that in (B-V) the colour is approximately constant at (B-V) = -0.1 $\pm$ 0.2 between 5 and 25 arc seconds from the star, this being slightly bluer than V380 itself, while there is an indication of reddening beyond 25 arc seconds. The (U-B) index was not so uniform, but gave a mean value of (U-B) = -0.5 within 25 arc seconds, with indications of blueing at larger offset distances ((U-B) at 30" = 1.0). Brück also measured (V-Hα), deriving a value of 2.8, and (V-I), finding 0.9, but for the central region only. She found that the nebular brightness fell off most rapidly to the NE of V380, and least rapidly to the W beyond the dark obscuration. The extinction produced in the dark obscuration, although too faint for her measurements, was estimated to exceed 4 magnitudes.

Herbig (1960) has argued on several grounds that the distance to NGC1999 must correspond to the distance of the Orion cloud (470 pc.):

i) The Hβ and λ3727 emission lines of the field nebulosity in this region can be detected overrunning the dark obscuration, thus indicating that the nebula is not a foreground object.

ii) V380 cannot be too far imbedded in the dark cloud, or the reddening observed would be much larger.

iii) Using an absolute magnitude based on a spectral type of A1, the distance modulus is in close agreement with that of the Orion cloud (See also the discussion in section 10.2.3).

iv) The close agreement between the emission line velocities of V380 and the HI velocity in the cloud may indicate an association.

Using this distance of 470pc., the linear extent of the nebula is ~0.2pc., and the dark obscuration has a radius of ~0.013pc. Brück has suggested that the obscuration may be similar to the globules studied in the Orion nebula by Penston (1969), and a comparison with NGC2261, associated with the star R Mon. (studied in the same paper), leads her to speculate that NGC1999 may be a cometary type nebula seen head-on.

10.2.3 Discussion of Previous Photometry

Brück (1974) has summarised previous measurements of the brightness and colour of V380 Orionis and discussed the results in the light of the contribution from the inner nebulosity of NGC1999 when using large apertures in photo-electric work.

Herbig (1960) quotes V = 9.7 to 10.1, with (B-V) = 0.34, while Mendoza (1968) gives V = 10.26, (B-V) = 0.44. Some discrepancy might be expected due to intrinsic variability of the star, but Brück has shown that the

inner brightness profile of the nebula is inconsistent with a star brighter than V = 10.5, B = 10.85, and attributes the brighter estimate to the inclusion of the bright, bluer, inner nebulosity. Her calculations show that inclusion of the nebulosity within the 36″ diaphragm used by Mendoza gives values in agreement with his results, indicating that no variability need be invoked to explain these differences.

On the basis of her corrected magnitudes, Brück finds (B-V) = 0.25, and deduces a reddening E_{B-V} = 0.22, based on an intrinsic (B-V) of 0.03 appropriate to spectral type A1. Assuming a visual extinction $A_V \approx 1$ magnitude to the Orion cloud, this reddening is fully accounted for without requiring V380 to be imbedded in the cloud. If we use a ratio of total to selective extinction ($R = A_V/E_{B-V}$) of 3.3, use E_{B-V} = 0.22 and an absolute magnitude M_V = + 0.8 appropriate to spectral type A1, the expected visual magnitude is V = 9.9 at a distance of 470pc., contrary to Brück's claim of agreement with the measured V = 10.6.

While Brück's measurements of B and V magnitudes are undoubtedly more accurate than previous large aperture photo-electric measurements, inspection of the spectro-photometry by Garrison and Anderson (1978b) shows that her estimate of the reddening may not be reliable. A peak at B wavelengths of ~0.2 magnitudes above the smooth continuum, attributable to emission lines, may make the

assumption of an intrinsic (B-V) for the star of 0.03 unjustified.

Garrison and Anderson (1978b) derive the reddening by comparison with a theoretical model over a wide wavelength range, and although they use a slightly earlier spectral type (B8-A0), obtain good agreement with theory, using an Orion-type reddening law corresponding to $E_{B-V} \sim 0.5$. The question of the reddening cannot, however, be considered settled, as the inclusion of the bright inner nebulosity by Garrison and Anderson cannot be discounted in the light of Brück's calculations. If we assume a reddening of ~ 0.3 magnitudes due to interstellar extinction to the Orion cloud, and attribute the remaining 0.2 magnitudes of E_{B-V} to the dust of the cloud itself (and use R = 5 within the cloud), we find, taking an absolute magnitude $M_V = +0.34$ for type B9, an expected visual magnitude of 10.61, agreeing very well with Brück's measured value.

In view of the uncertainties involved, however, not too much significance can be placed on this agreement produced by Garrison and Anderson's reddening determination. This is unfortunate, since the reddening may provide a means of placing constraints on possible dust-scattering models of NGC1999 to be described later. However, the values V = 10.6, B = 10.85, $E_{B-V} = 0.5$ will be used, when needed in formulating these models, with the corresponding uncertainties well in mind.

10.3 The New Observations of NGC1999

10.3.1 Observations

New polarimetric observations of NGC1999 were made during January 1978 by the author, Dr. S. M. Scarrott (University of Durham), Dr. P. Murdin (Anglo-Australian Observatory) and Dr. K. Taylor (University of New South Wales), using the 3.9 metre Anglo-Australian Telescope at Siding Spring, NSW, Australia.

The Durham polarimeter and McMullan electronographic camera were used at the f-15 Cassegrain focus with a broad band visual filter (GG455 + BG38) covering the wavelength range 0.45μm. to 0.65μm. (Figure 13.11) and 8 exposures, each of 15 minutes, were made to produce a complete polarisation map of the object. During the observations the sky was moonless, but some changing atmospheric "seeing" and extinction was noted.

Figure 10.1 shows a mosaic picture of NGC1999 composed of photographs taken later of two polarimeter electronographs. It can be seen that the faint outer parts of the nebulosity extend over a large part of the polarimeter field.

10.3.2 Data Reduction

The electronographs obtained at the AAT were digitised at the Royal Greenwich Observatory and the data was reduced using the technique described in part 1 of this thesis. The resulting polarisation map is shown

in figure 10.2 where the integration areas are 3.3 arc seconds square. The sky signal was sampled from the extreme eastern edge of the field.

Due to the lack of stars in this region of the Orion dark cloud, image alignment has been accomplished using the polarimeter grid spots alone (5.3) but, because of the extremely accurate tracking of the AAT, the mis-alignment produced by this lack of stars is unlikely to exceed 1 pixel. Since the relative alignment of the left and right hand polarimeter images does not, in any case, require field stars, errors from this cause do not affect the measurement of polarised quantities. As a check, the data reduction has been repeated with several images arbitrarily displaced by up to 4 pixels - no significant change in the polarisation map resulted.

Mutual alignment of the "grids IN" and "grids OUT" halves of the map has been accomplished by finding a location on each half which represents the "centre" of the polarisation pattern, according to the criterion that:

$$\int = \Sigma \frac{(\theta_m - \theta_T)^2}{\sigma^2} \qquad 10.2$$

be minimised. (Here θ_m is the measured polarisation orientation, with error σ, and θ_T is the expected orientation assuming the pattern to be circular). Shifts in the origins of the two halves were then made to make these 2 centres co-incide.

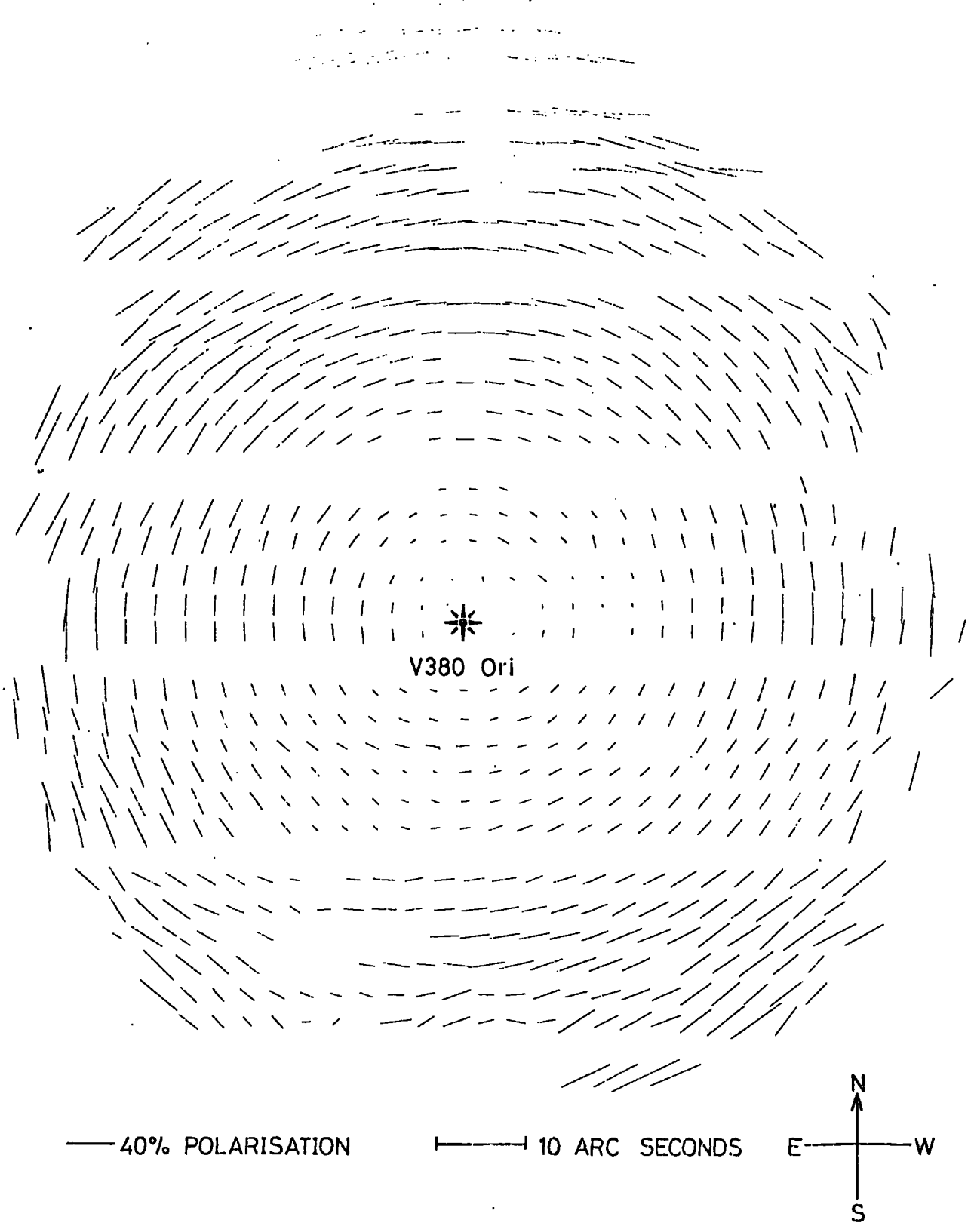


Fig.10.2 The polarisation map of NGC1999

In the course of this alignment procedure, the polarisation orientations were found to be perfectly circular, with a centre within the brightest part of the image of V380 Ori. Attempts to make correction for the interstellar polarisation reported by Garrison and Anderson (10.2.1), however, introduced noticable ellipticity into the pattern. It was thus concluded that no interstellar correction was necessary, and the value based on observations of V380 Ori. by Garrison and Anderson must presumably be ascribed to the inclusion of part of the highly polarised nebula within their measurement aperture.

For the reasons outlined in section 7.4.4, data validation has not been carried out on the central regions of the nebula, within 15 arc seconds of the star. The measurements plotted correspond to those regions of nebula with a recorded brightness above 3 PDS units ($\sim$23 mag./sq. arc sec.) and the central image of V380 Ori. is over-exposed.

10.3.3 Features of the New Data

The most readily apparent feature of the new polarisation map is the striking circular pattern centred on V380 Ori. As was noted above, this pattern is truely circular and is consistent with a single illuminating source for the nebula, co-incident with the optical image of V380 - polarisation being produced

by light scattering in the nebular medium and resulting in a pattern with the electric vector everywhere perpendicular to the scattering plane.

Since the nebular spectrum contains only lines found in V380 itself and the nebular colour is similar to that of the star, the presence of free-electron scattering can be ruled out, and NGC1999 appears to be a simple reflection nebula seen by the light scattered by nebular dust.

Figure 10.3 shows an image of NGC1999 produced from the processed polarisation data, using a grey-scale devised by Dr. W. S. Pallister of the University of Durham for use on a teletype computer terminal. The full dynamic range of the electronographic technique is here represented in a density range which can be appreciated by eye, the grid overlap regions being interpolated to provide a continuous visual image. The dense obscuration which is so obvious on short exposure photographs can be seen to affect only the central regions of the nebula, the outer regions being largely featureless with approximately circular contours.

Using the same technique, figure 10.4 shows an image of the polarised intensity, and figure 10.5 shows the % polarisation within a circle of radius 50" (Beyond this distance the data is unreliable). The highest polarisations recorded are $\sim$40%. It is clear from these images that the dense obscuration has no significant

effect on the measured polarisations, a conclusion which is confirmed by figures 10.6 and 10.7 showing traces of total intensity and % polarisation in a direction W-E through the star and obscuration. The measurements in the obscuration are affected only by an increased error due to the lower surface brightness, and it must thus be concluded that the obscuration is foreground, largely unassociated with the nebula and without significant polarising properties. This conclusion is also supported by the lack of any evidence of edge-brightening, which might be expected if the obscuration were a dense region within the nebula.

The polarisations measured show a general trend, rising from about 10% at 15" from the star to about 30% on the periphery of the data. The rise, however, is not the same in all directions; as can be seen in figure 10.5, the NE quadrant shows somewhat higher polarisations than elsewhere and this appears to define a direction of asymmetry in the nebula, running SW-NE through the star at PA $\sim 40^{\circ}$.

The variation of surface brightness with offset distance is shown in figures 10.8 (which shows a trace through the star running NW-SE at PA = 130°) and 10.10 (running SW-NE at PA = 40°). Also shown, in figures 10.9 and 10.11 are the corresponding polarisation measurements lying within "cones" centred on these directions and subtending an angle of 10° at the star.

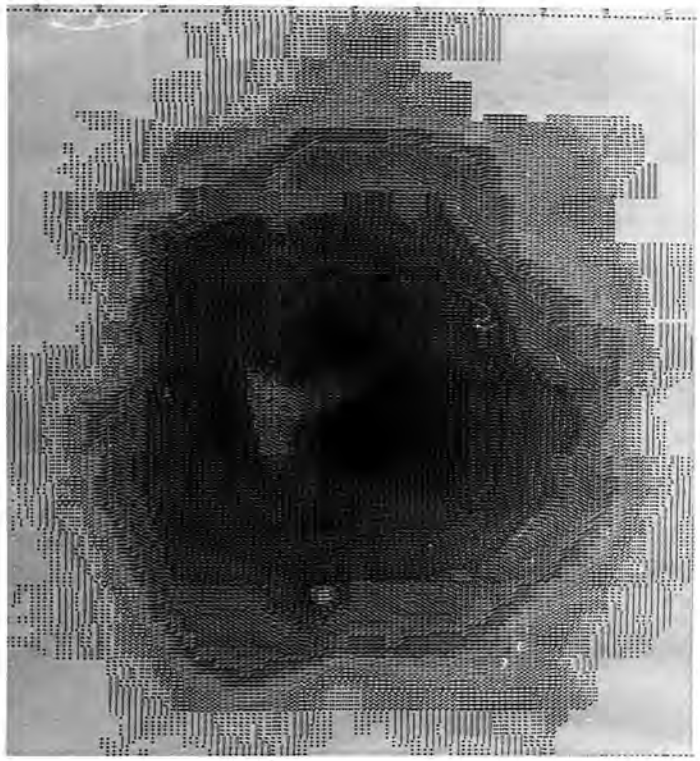


Fig. 10.3 Printed image of total intensity in NGC1999

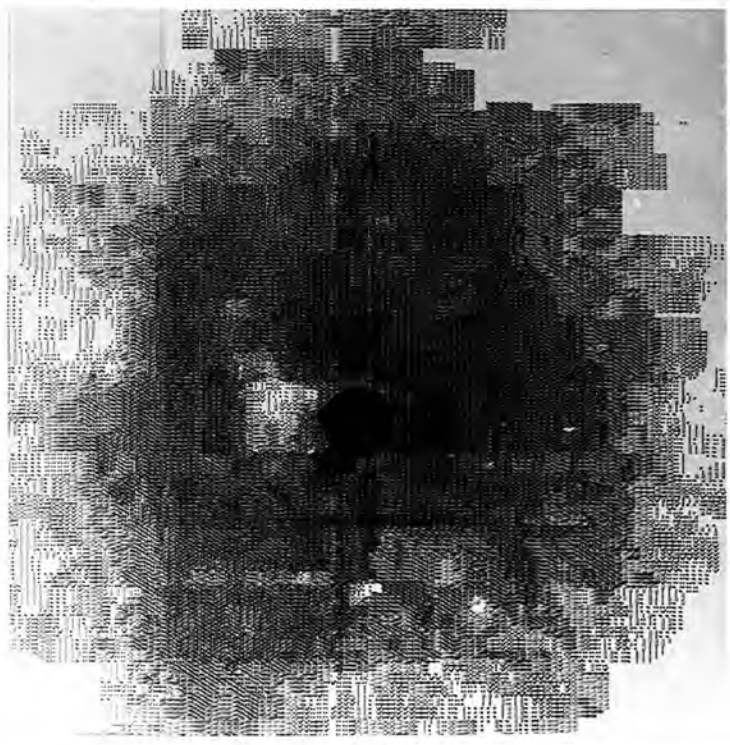
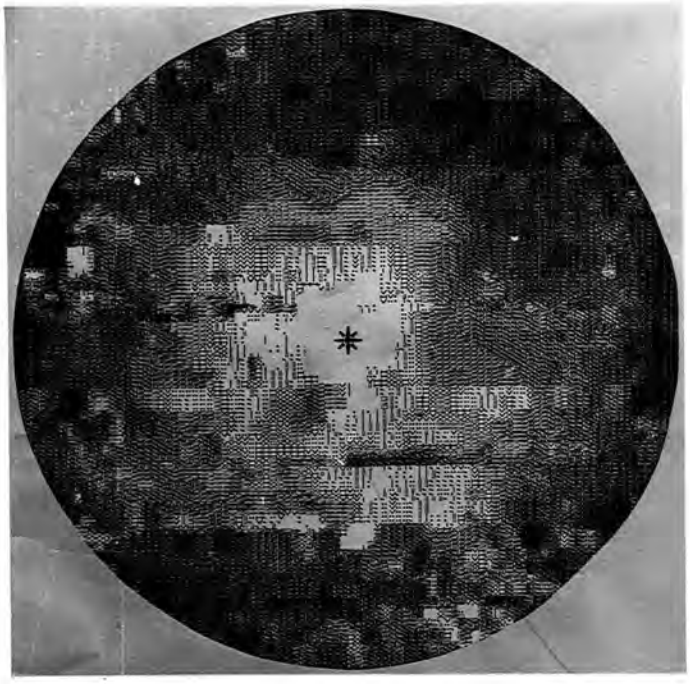


Fig. 10.4 Printed image of polarised intensity in NGC1999



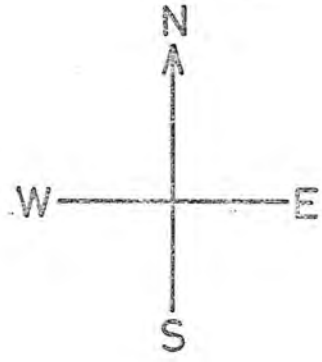


Fig.10.5 Printed image of % polarisation in NGC1999

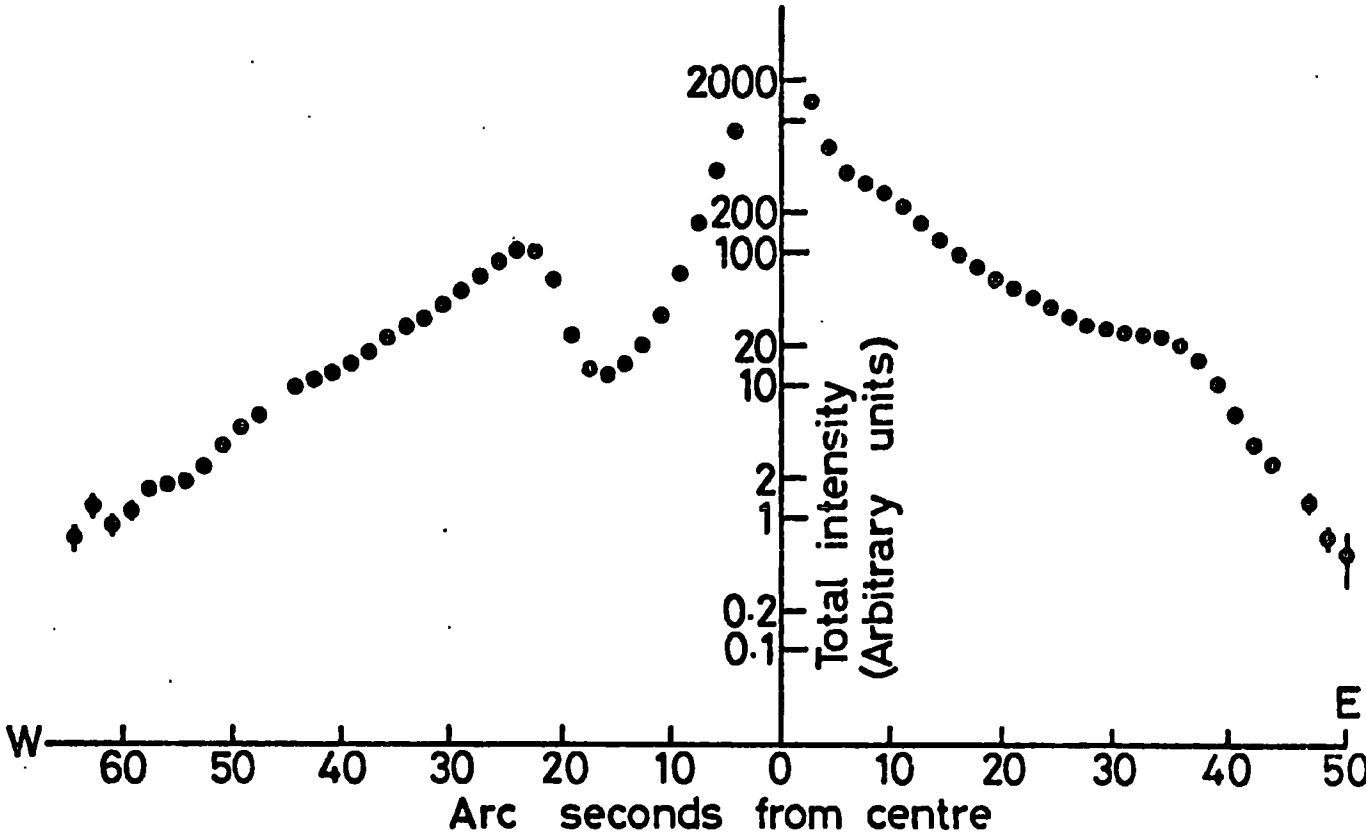


Fig.10·6 Total intensity through the star and obscuration

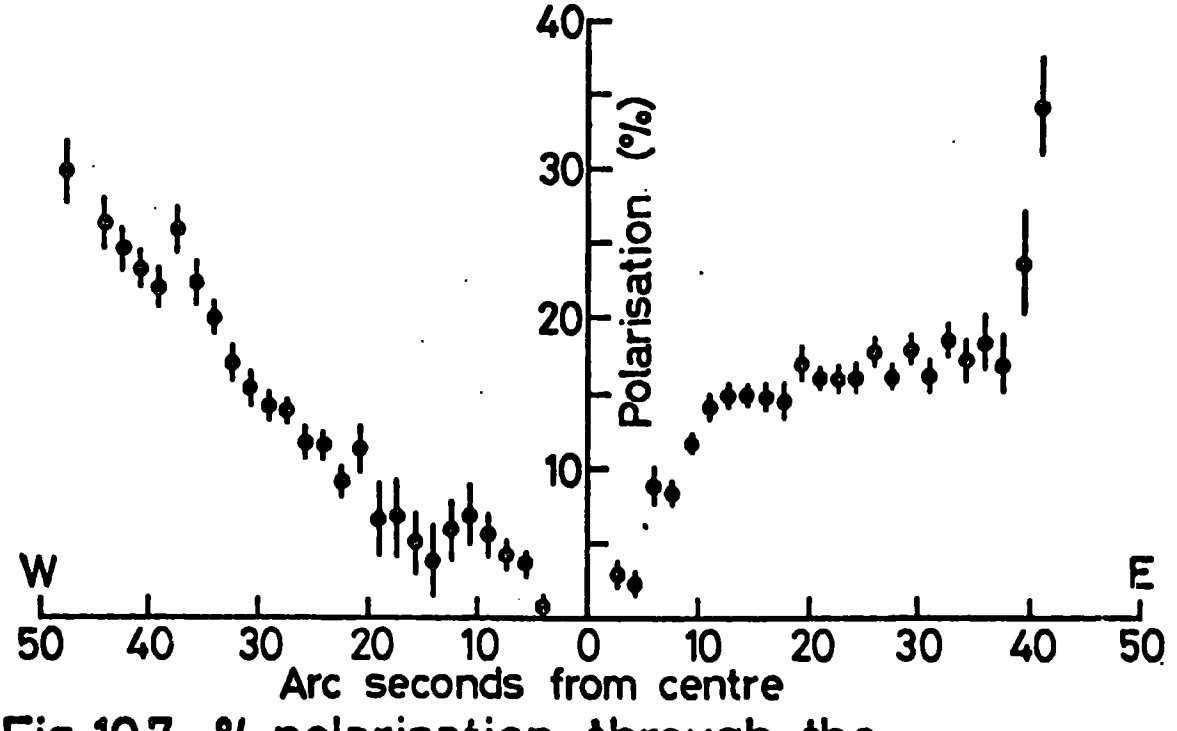


Fig.10·7 % polarisation through the star and obscuration

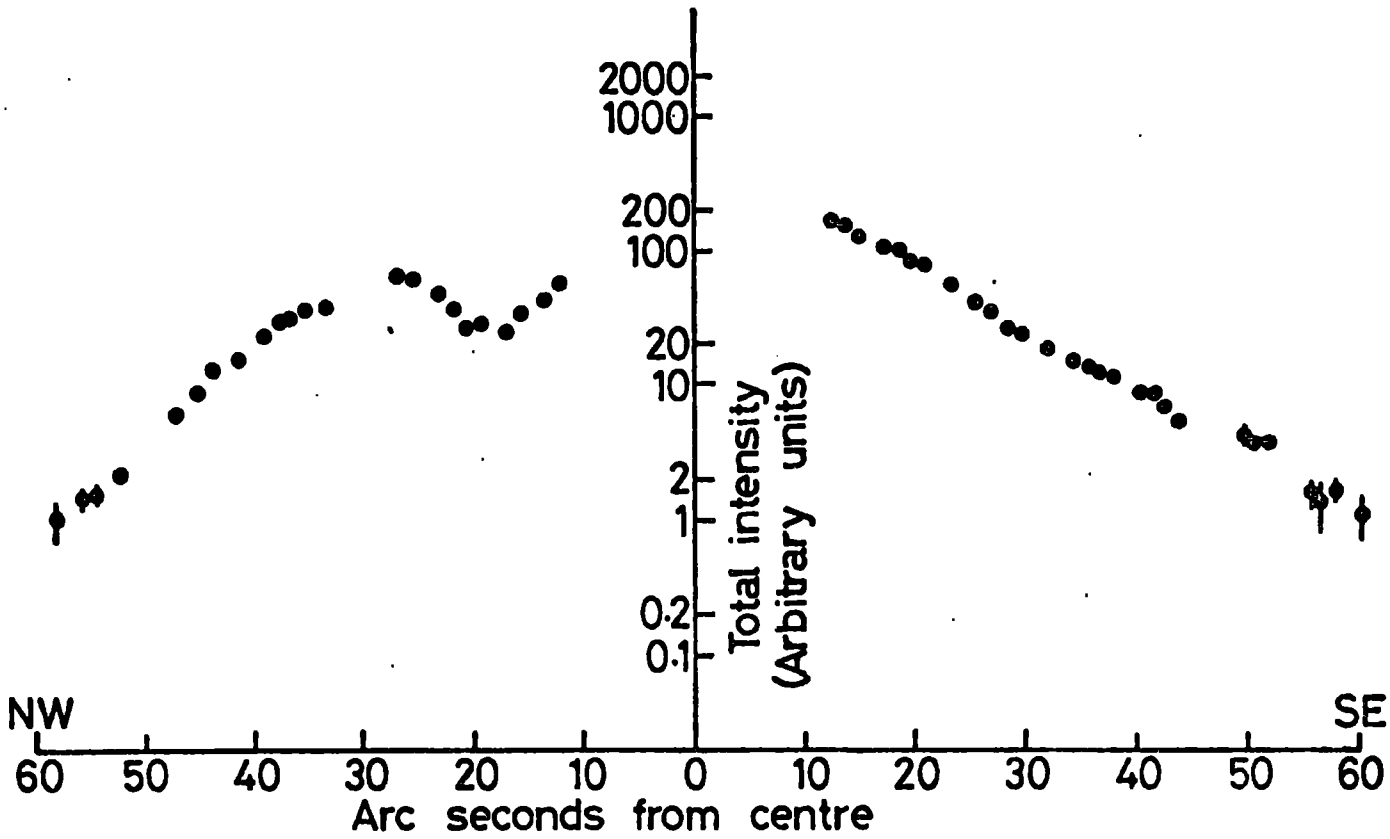


Fig.10·8 Total intensity NW-SE

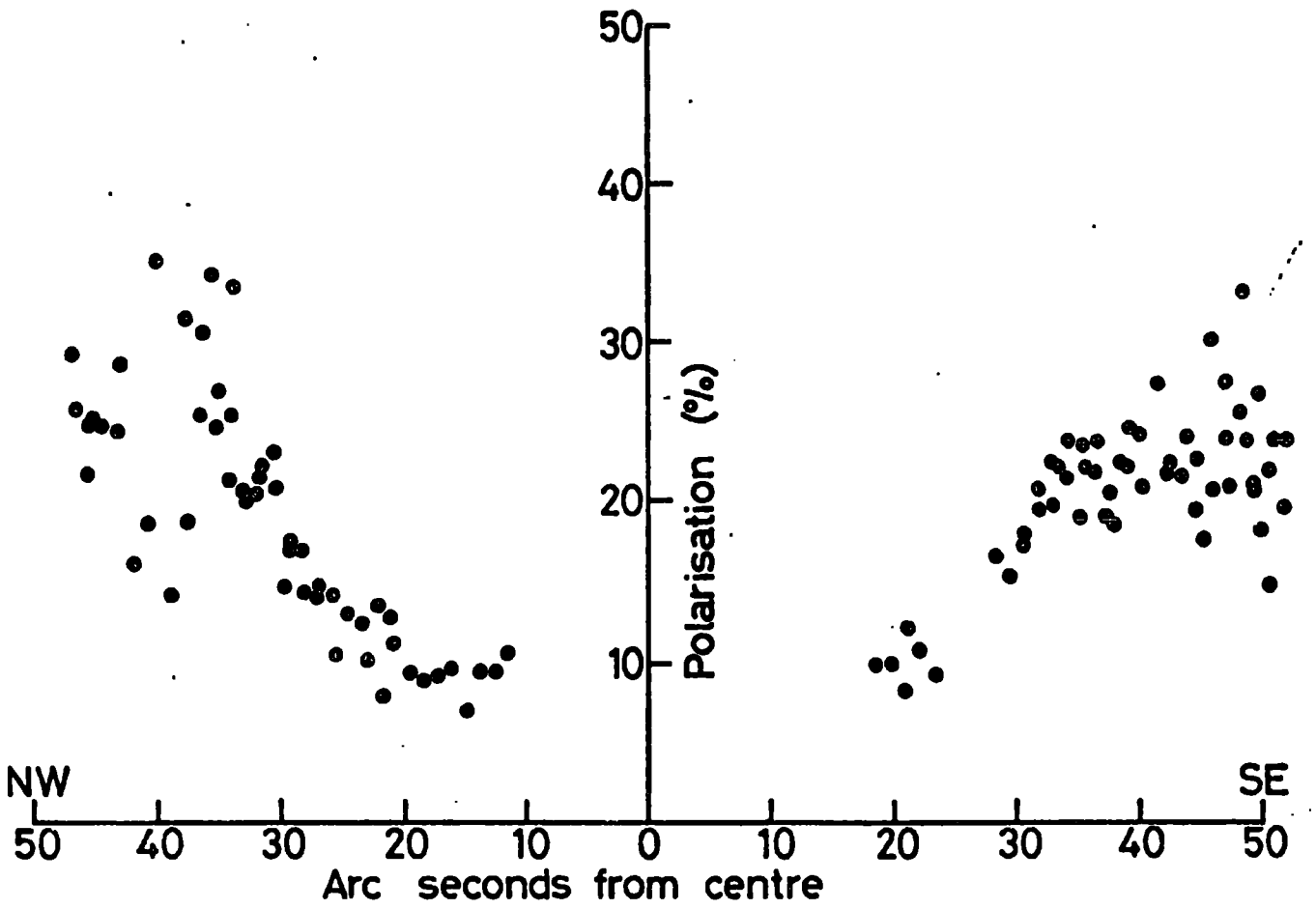


Fig.10·9 % polarisation NW-SE

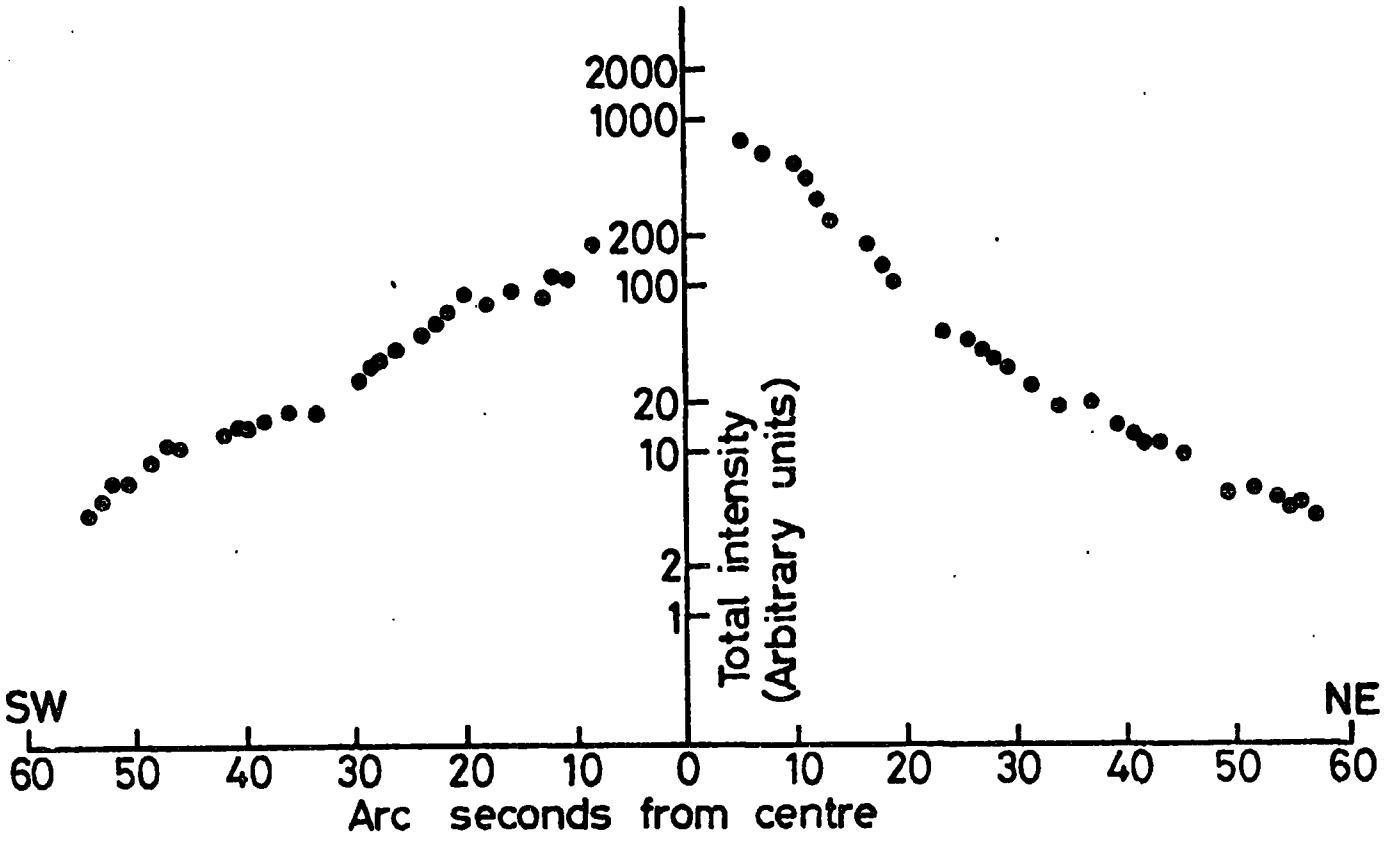


Fig.10·10 Total intensity SW-NE

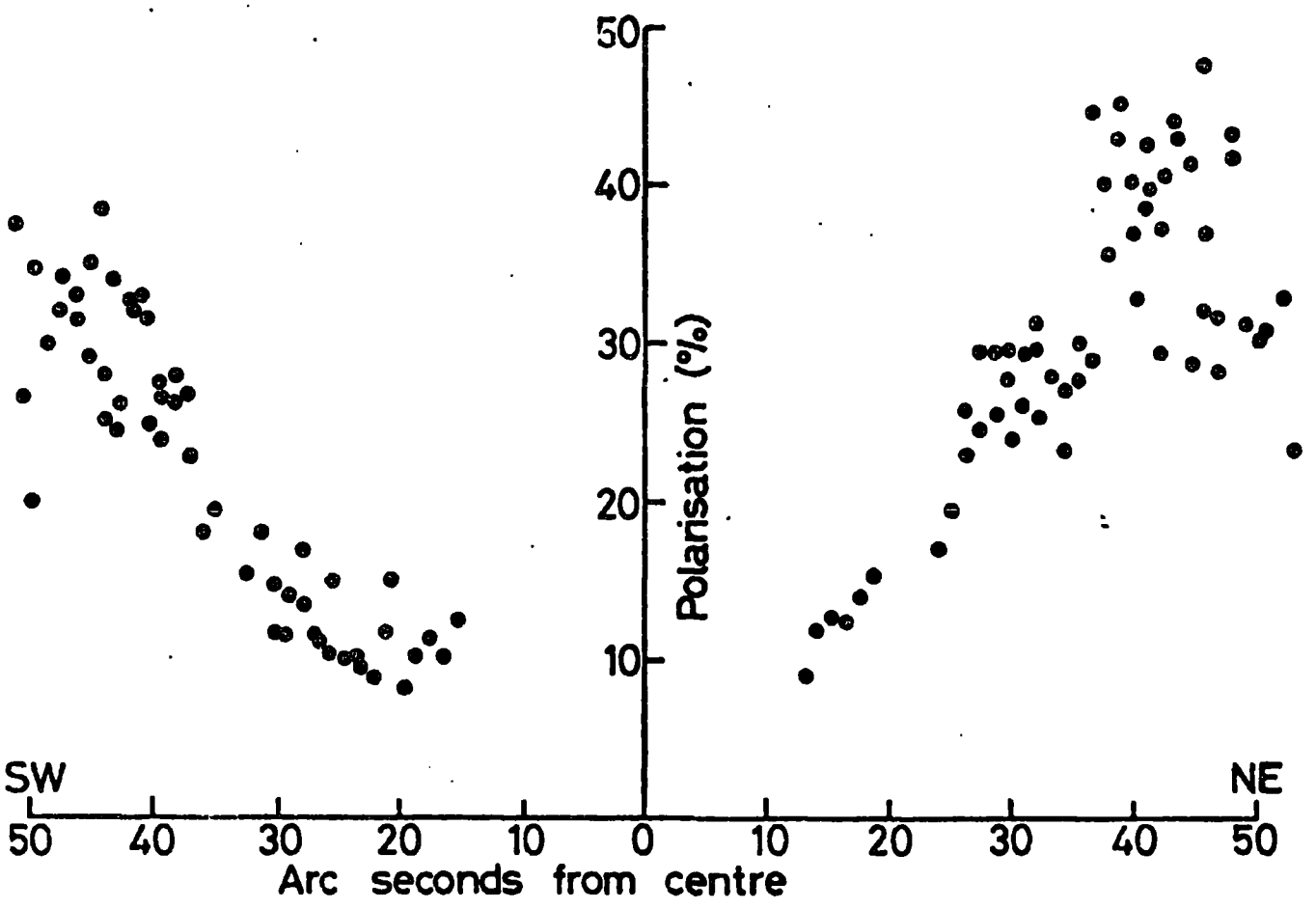


Fig.10·11 % polarisation SW-NE

It can be seen that the surface brightness falls off approximately exponentially with distance in all directions beyond the central obscuration. Inspection of figure 10.3, however, shows that the fall-off is slightly more rapid to the NE than elsewhere, a fact also noted by Brück (1974).

The polarisation measurements show a marked asymmetry; the polarisation rising to ~40% on the NE periphery of the data, compared with only 25 to 30% elsewhere. At all offset distances from the star the NE quadrant shows predominantly higher polarisations.

10.4 Conclusions

The above observations indicate that, although the dark obscuration near the centre of the nebula confuses the optical image, most of the nebular surface is unaffected and NGC1999 is probably a relatively simple reflection nebula associated with the dark cloud which contains the Orion Nebula.

The marked asymmetry in the polarisation measurements requires some explanation. An effect of this size cannot be instrumental in origin, nor is it caused by image mis-alignment, since deliberate displacement of the polarimeter images does not affect the measurements (10.3.2) - it therefore appears to be real. Since it has no special geometrical link to the dark obscuration,

this is unlikely to provide the key and uncorrected interstellar polarisation is probably too small an effect and would, in any case, distort the circular pattern of polarisation observed. Similarly, magnetic alignment of non-spherical dust grains within the nebula is likely to be too small an effect, may distort the centro-symmetric pattern (as, for example, in the Merope nebula - Elvius and Hall, 1966) and must in some way be confined to one quadrant, since no enhancement of polarisation is seen to the SW as might otherwise be expected. The most natural explanation appears to be some form of tilt with respect to the plane of the sky (or a similar geometrical asymmetry). This may explain the fact that the region of enhanced polarisation subtends approximately the same angle at the star at all offset distances and may also tie in with the slightly faster brightness fall-off to the NE.

Using the measurements of Brück (1974) to normalise the brightness to B = 19.8 magnitudes per sq. arc sec. at 20" from V380, integration over the exponential fall-off shows that the light received from the nebula outside 10" from the star is of the same order as from the star itself (B = 10.85) in this waveband. Consequently, the nebula must subtend an appreciable solid angle at the star in order to intercept this amount of light. The star is thus likely to be contained within the nebula and optical depths of order unity may be present. As

a result, stellar extinction by the nebular dust is likely to be present, although the reddening data suggest that this is unlikely to amount to 1 magnitude at visual wavelengths.

Although many small variations in the % polarisation are seen (Figure 10.5), some of which may be real and some of which are attributable to data errors resulting from emulsion non-uniformity, the overall structure of the data is sufficiently well defined to attempt to represent it by using an idealised nebular model with a simple geometry. The remaining chapters of this thesis describe an attempt to do this, while at the same time investigating those properties of dust grains which relate to polarimetric observations such as these.

CHAPTER 11

LIGHT SCATTERING FROM DUST GRAINS

11.1 Introduction

Scattering of light from generally shaped particles presents a very difficult theoretical problem, requiring the solution of Maxwell's equations of electromagnetic wave propagation, both inside and outside the particles, with appropriate boundary conditions applying on the surface. Because of this difficulty, only a small number of simplified cases have been treated theoretically - the cases of infinitely small particles (Rayleigh scattering) and infinitely large particles (Geometric scattering) being the first to be investigated. The first rigorous solution for particles of arbitrary size was obtained by Mie (1908) for the case of smooth, homogeneous spheres.

Since this time, a combination of exact and approximate methods has been used to derive results for cylinders, ellipsoids and a number of other simple geometric shapes (A comprehensive description of scattering by a wide range of particles is given by van de Hulst, 1957). Many of these treatments, however, have concentrated on calculating the total cross sections of the particles, rather than the intensity and state of polarisation of the light scattered in a particular direction (The differential cross section). Consequently,

the polarisation induced when light is scattered from particles other than spheres is still poorly understood.

While it is known from the existence of interstellar polarisation, induced by extinction, that real interstellar dust particles must have a degree of elongation, insufficient information is available to construct particle-shape models and, because of this and the theoretical difficulty involved in treating non-spherical particles, most work on the optical properties of interstellar dust has used Mie's theory as a basis. Several workers have investigated the validity of this approximation by microwave analogue experiments (A summary of this work is given by Zerull and Giese, 1974) and the conclusions can be briefly summarised by saying that so long as the deviation from sphericity is not large, good agreement with the overall structure of Mie's theory is obtained, although the fine detail in the scattering functions is not reproduced. Since astronomical applications are concerned only with the overall structure of the scattering functions, the approximation would seem justified.

11.2 The Geometry of Scattering

The geometry of light scattering is shown in figure 11.1. A beam of light, having components polarised perpendicular and parallel to the scattering plane with complex amplitudes A_1 and A_2 (1.1) undergoes scattering

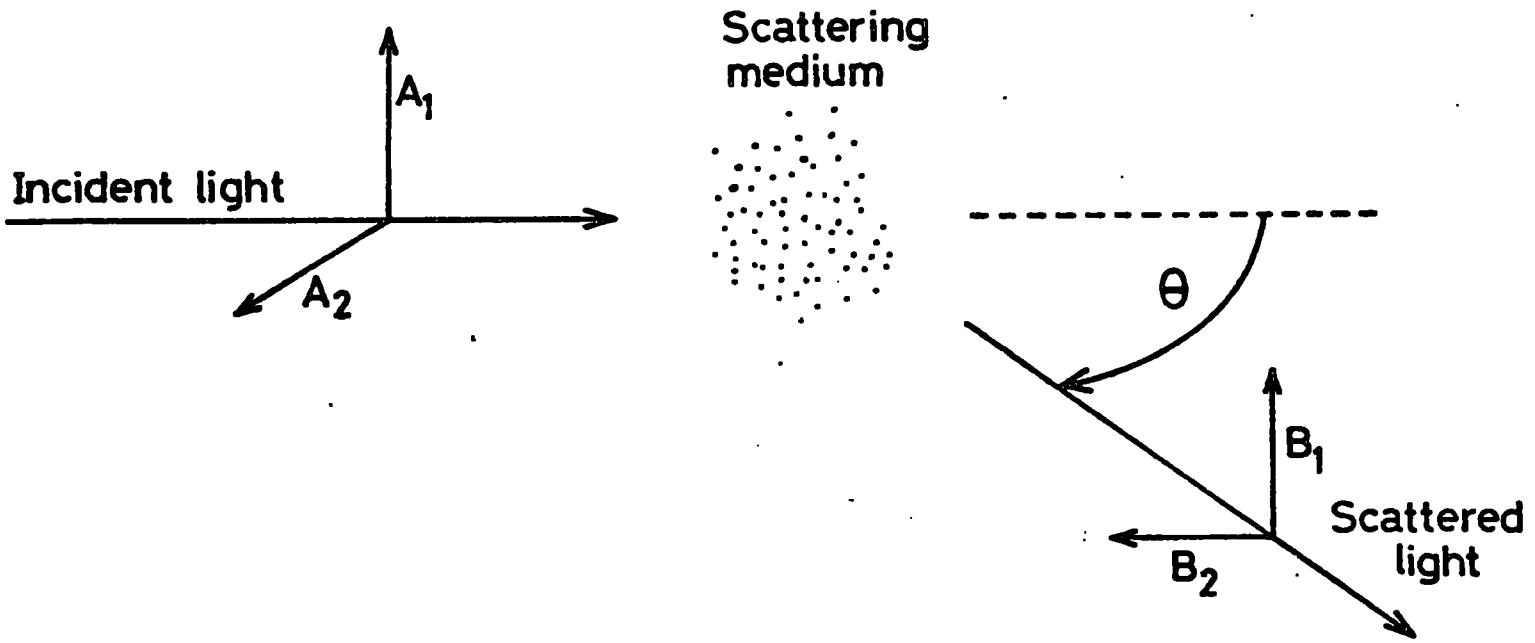


Fig.11.1 The geometry of light scattering

through an angle θ. In the scattering process, these complex amplitudes undergo changes in amplitude and phase and the resulting amplitudes B_1 and B_2 determine the new state of polarisation after scattering. In general, the amplitudes are related by a matrix equation:

$$(B_1, B_2) = \begin{bmatrix} S_1(\theta) & S_3(\theta) \\ S_4(\theta) & S_2(\theta) \end{bmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \qquad 11.1$$

and all 4 complex functions $S_i(\theta)$, $i = 1,4$ must be known to describe the scattering fully. In the case of scattering from spheres, however, symmetry requires that $S_3 = S_4 = 0$, so that the amplitudes perpendicular and parallel to the scattering plane may be considered independently and the scattering is fully described in terms of two complex scattering amplitudes $S_1(\theta)$, $S_2(\theta)$.

If the incident light is unpolarised, the two incident components are of equal amplitude and mutually independent, and the phase changes which take place on scattering are irrelevant. In such a case only the absolute values of S_1 and S_2 need be known, and the scattered light is always plane polarised (i.e. no circular or elliptical polarisation is produced). The <u>intensities</u> of the scattered components are then proportional to $|S_1|^2$ and $|S_2|^2$, and the percentage polarisation of the scattered light is:

$$P(\theta) = \frac{|S_1(\theta)|^2 - |S_2(\theta)|^2}{|S_1(\theta)|^2 + |S_2(\theta)|^2} \times 100 \qquad 11.2$$

while the total scattered intensity within a unit solid angle due to a unit incident light flux (i.e. the differential scattering cross section) is:

$$I(\theta) = |S_1(\theta)|^2 + |S_2(\theta)|^2 \qquad 11.3$$

11.3 The Mie Scattering Formulae

Mie's theory gives the scattering functions of a sphere in terms of a dimensionless size parameter:

$$X = \frac{2\pi a}{\lambda} \quad , \qquad 11.4$$

where a is the sphere radius and λ the wavelength of the light, and a complex refractive index for the sphere material:

$$m = \sqrt{K - 2i\sigma\lambda/c} \quad , \qquad 11.5$$

where K is the dielectric constant, σ the conductivity and c the velocity of light.

The total cross section of the sphere for removal of light from a beam by scattering is given by:

$$C_{SCA} = \pi a^2 \; Q_{SCA} \qquad 11.6$$

where the efficiency factor Q_{SCA} is given by the theory as:

$$Q_{SCA} = \frac{1}{X^2} \sum_{n=1}^{\infty} (2n+1) \left[|a_n|^2 + |b_n|^2 \right] \qquad 11.7$$

and the functions a_n, b_n are complex functions of m and X.

The cross section for removal of light from a beam by both absorption (resulting in heating of the sphere) and scattering has an efficiency factor:

$$Q_{EXT} = \frac{2}{X^2} \sum_{n=1}^{\infty} (2n+1) \; Re\left[a_n + b_n \right] \qquad 11.8$$

while the cross section for absorption alone has efficiency factor:

$$Q_{ABS} = Q_{EXT} - Q_{SCA} \qquad 11.9$$

Also given by the theory are the complex scattering amplitudes:

$$S_1(\theta) = \sum_{n=1}^{\infty} \frac{(2n+1)}{n(n+1)} \left[a_n \pi_n (\cos\theta) + b_n \tau_n (\cos\theta) \right] \qquad 11.10$$

and:

$$S_2(\theta) = \sum_{n=1}^{\infty} \frac{(2n+1)}{n(n+1)} \left[b_n \pi_n (\cos\theta) + a_n \tau_n (\cos\theta) \right] \qquad 11.11$$

where π_n and τ_n are functions derived from the Legendre polynomials.

The differential cross section of equation 11.3 is found in terms of s_1 and s_2 by setting:

$$S_j = \frac{1}{\sqrt{2}} \frac{\lambda}{2\pi} s_j \quad , \quad j = 1,2 \qquad 11.12$$

A summary of the important results from Mie's theory together with definitions of the functions a_n, b_n, π_n and τ_n are given by Wickramasinghe (1973), while van de Hulst (1957) gives a derivation of the results from first principles.

11.4 Evaluating the Mie Scattering Functions

A computational scheme for evaluating the Mie scattering functions is described by Wickramasinghe (1973), based on the evaluation of the functions a_n, b_n, π_n and τ_n from their recurrence relations. The summations of equations 11.7, 11.8, 11.10 and 11.11 can be approximated in practice by a finite number of terms which does not exceed 20 for $X \lesssim 15$ and an accuracy of $\sim 0.1\%$. Wickramasinghe gives a table showing the number of terms required for various values of m and X.

This computational scheme has been implemented to evaluate the cross sections and scattering amplitudes of the previous section and the accuracy checked by comparison with Wickramasinghe's many tabulated values over a wide range of all variables.

11.5 The Form of the Mie Scattering Functions

The scattering functions of a single sphere, calculated from Mie's theory, show a wide diversity of forms depending on the size and refractive index of the sphere. For spheres with radii much less than the wavelength of the light, and with refractive indices near unity, the form may be relatively smooth with variations in θ or X. However, for sizes comparable to, or larger than, the wavelength and for refractive indices differing significantly from unity, the form may be very irregular, showing numerous oscillations with θ and rapid alternation between positive polarisation (perpendicular to the scattering plane, with $|s_1| > |s_2|$) and negative polarisation (parallel to the scattering plane, with $|s_1| < |s_2|$). Examples of $|s_1|$ and $|s_2|$ for two sphere sizes are shown in figure 11.2 for a refractive index m = 1.63 - 0.05i, possibly typical of interstellar silicate material.

The oscillations are attributable to resonances within the sphere, occurring at different scattering angles depending on the exact size and refractive index. As might be expected, the effect of these resonances is

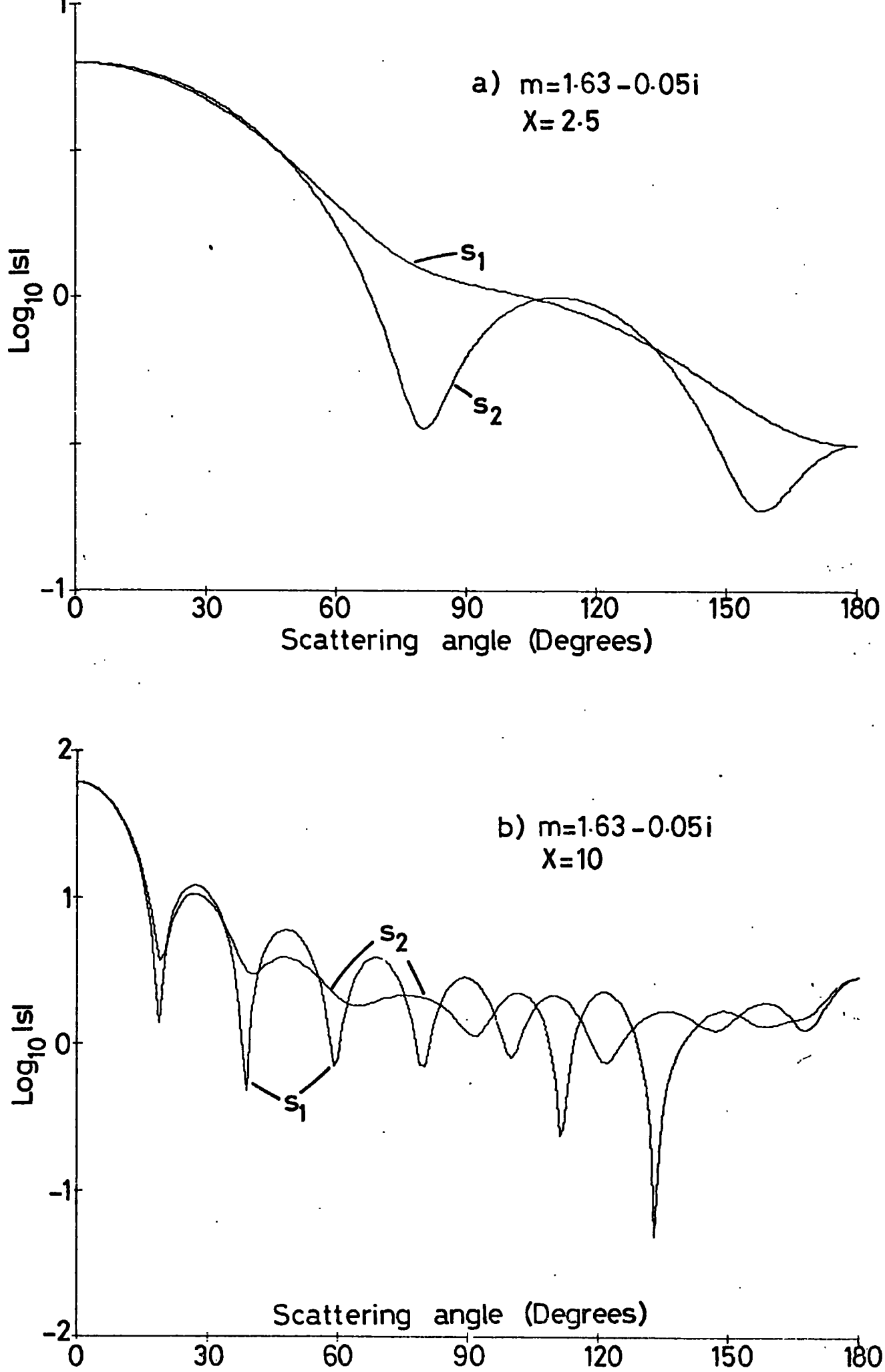


Fig.11·2 Examples of Mie scattering amplitudes for single spheres

damped as the complex (absorptive) part of the refractive index is increased, although the shape of the scattering functions may still be very irregular. Further graphical examples of these functions are given by Wickramasinghe (1973).

11.6 Integration Over a Size Distribution of Spheres

To use Mie's scattering theory to represent the scattering of light by interstellar dust grains, we must average the scattering functions for a single sized sphere over the range of sizes found amongst interstellar dust grains.

To perform this averaging, we assume the dust grains to have a range of sizes with a distribution function n(a), such that the probability of a given grain having a radius between a and a + da is n(a) da. This size distribution function must be normalised to have:

$$\int_0^\infty n(a)\,da = 1 \qquad 11.13$$

By performing an integration over the amount of light scattered by spheres of each size, we can find average scattering cross sections for the collection of grains:

$$G(m) = \int_0^\infty C(m,X).n\,(a).da = \int_0^\infty \pi a^2 Q(m,X).n\,(a).da \qquad 11.14$$

and average scattering functions:

$$F_j(m,\Theta) = \int_0^\infty |s_j(m,X,\Theta)|^2 . n(a).da \quad , \; j = 1,2 \qquad 11.15$$

where X is given by equation 11.4. In this way, average optical properties are ascribed to a particular size distribution of particles and it is usual to refer to these average properties as the scattering functions of an imaginary "average grain".

Difficulty has been experienced in performing these integrations by previous workers (Greenberg and Hanner, 1970) due to the very irregular form of the Mie functions, particularly when the scattering angle is close to 180^o, the particle sizes large and the refraction index large and real. Simple integration methods using a fixed interval must allow for these "worst case" conditions by using a small step length, resulting in very inefficient integration in the cases when the functions in fact turn out to be relatively simple. Because this integration dominates the computing time requirements when performing scattering calculations, an adaptive method has been adopted. The integration range is split into sub-ranges, corresponding to intervals of 0.5 in X, and each sub-range is integrated with a routine from the NAG Fortran subroutine library which uses a method

described by Patterson (1968a,b) based on Gauss quadrature, with between 7 and 255 function samples optimally chosen to minimise the number of function evaluations needed. In practise, the number of evaluations can change by a factor of 10 according to the parameters of the scattering particles, so an adaptive approach such as this has considerable advantages.

The results of these integrations have been checked against the (corrected) values of Greenberg and Hanner and show perfect agreement.

11.7 Choice of Size Distribution Function

Very little information is available about the type of grain size distribution appropriate to interstellar dust. From the striking constancy of the observed interstellar reddening law, many have concluded that the optical properties of the interstellar dust are largely homogeneous throughout the galaxy and that the size distribution thus represents some form of steady state, or equilibrium distribution in the dust formation/destruction process. Considering an equilibrium between grain growth and a destruction factor leads to an equation satisfied by the size distribution function:

$$\frac{da}{dt} \cdot \frac{d\left[n(a)\right]}{da} + D(a) \cdot n(a) = 0 \qquad 11.16$$

where $\frac{da}{dt}$ is the grain growth rate (assumed constant, due to accretion) and D(a) is the destruction probability per unit time (Greenberg, 1966). Setting D(a) proportional to the grain cross sectional area results in a solution of the form:

$$n(a) \propto \exp\left[-5\,\frac{a^3}{a_o^3}\right] \qquad 11.17$$

(where the factor 5 is conventionally included as it makes a_o approximately equal to the mean grain size). This function gives a good approximation to an earlier form proposed by Oort and van de Hulst (1946) based on a grain growth and destruction model.

While this form has been used with some success in interpreting the observed interstellar reddening law, so too have a number of other functions, most notably the exponential:

$$n(a) \propto \exp\left[-a/a_o\right], \qquad 11.18$$

the Gaussian:

$$n(a) \propto \exp\left[-a^2/a_o^2\right] \qquad 11.19$$

and the form:

$$n(a) \propto a^{3/2} \exp\left[-\frac{1}{2}\frac{a^3}{a_*^3}\right] \qquad 11.20$$

Single grain sizes have also been used (due presumably to their simplicity) but are unlikely to be physically realistic. A survey of the success of a variety of dust grain models and size distributions in explaining the observed reddening law has been given by Wickramasinghe and Nandy (1972). It appears that the interstellar reddening law contains little information which can be used to define the size distribution.

Recently, Hong and Greenberg (1978) have shown that the observed maximum in the wavelength dependence of interstellar polarisation implied a mean grain size which differs slightly from that implied by the interstellar reddening law and they conclude that the form of equation 11.17 can most easily explain both sets of observations for dielectric grains (m real).

11.8 Effect of Size Distribution on Scattering Functions

Since there is undoubtedly still much uncertainty over the size distribution function of interstellar dust, the effects which the use of an inappropriate function might have on the calculated scattering functions has been investigated. It was found that for all the size

functions of section 11.7 (except the single size) the scattering functions could be made almost identical by a suitable choice of the parameter a_o.

Closer investigation shows that this is due to the predominant contribution which a relatively narrow range of grain sizes makes to the scattering functions. To illustrate this, figure 11.3 shows size functions of the forms of equations 11.17, 11.20 and 11.18, each adjusted to give a polarisation of 50% for scattering through 90°, using a refractive index m = 1.63 - 0.05i and a wavelength of 0.55μm. Although the size functions are dissimilar, the total light scattered as a function of grain size (obtained by weighting these curves with the scattering cross section appropriate to the grain size - figure 11.4) is remarkably similar in each case. In particular, the mean scattering size, defined by:

$$\langle a \rangle_{SCA} = \frac{\int_0^\infty a.C_{SCA}(a).n(a).da}{\int_0^\infty C_{SCA}(a).n(a).da} \qquad 11.21$$

is close to 0.15μm. for all 3 functions and it is clear that it is this parameter which primarily determines the form of the resulting scattering functions.

It is clear, also, from the positions of the curves

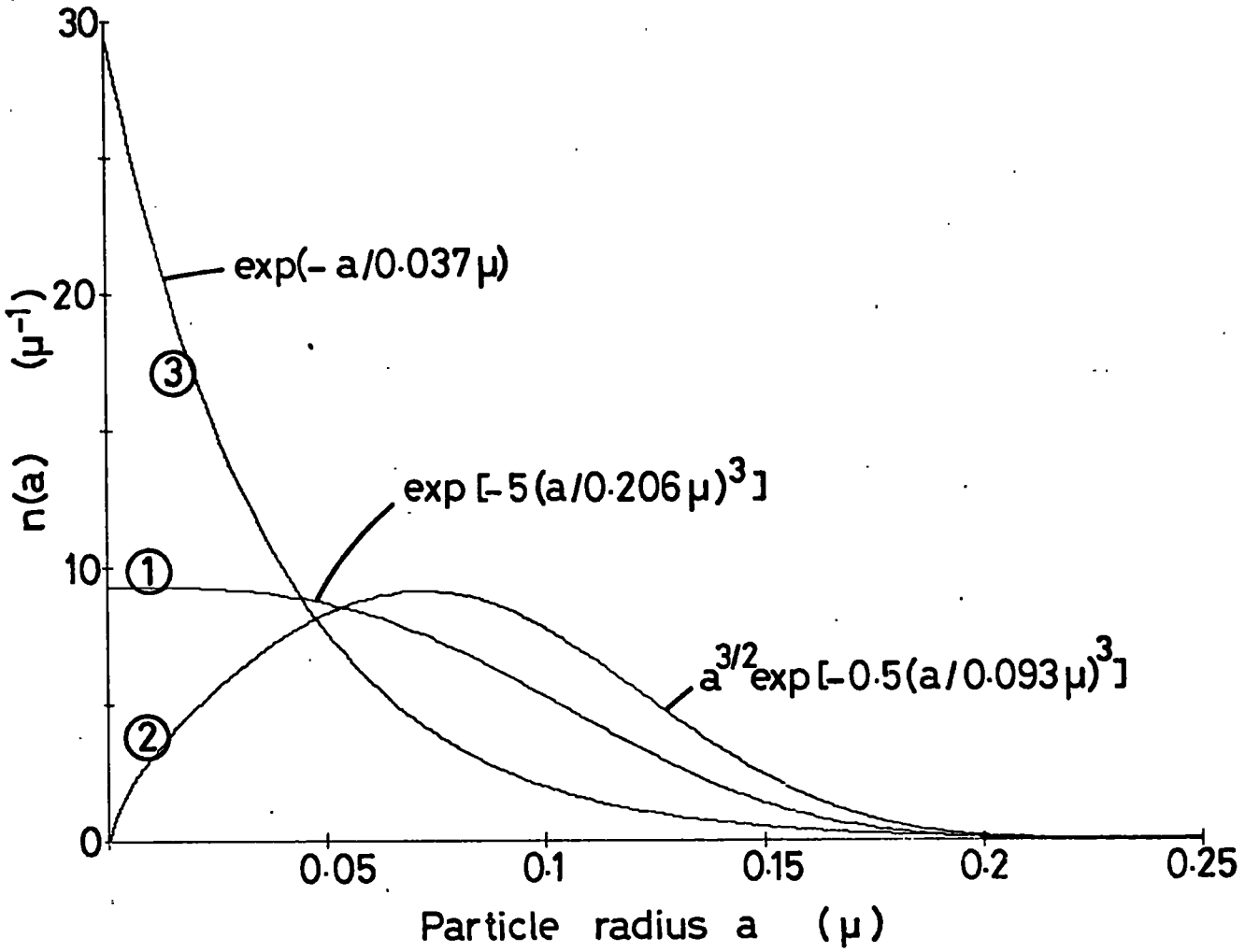


Fig.11·3 Normalised size distributions

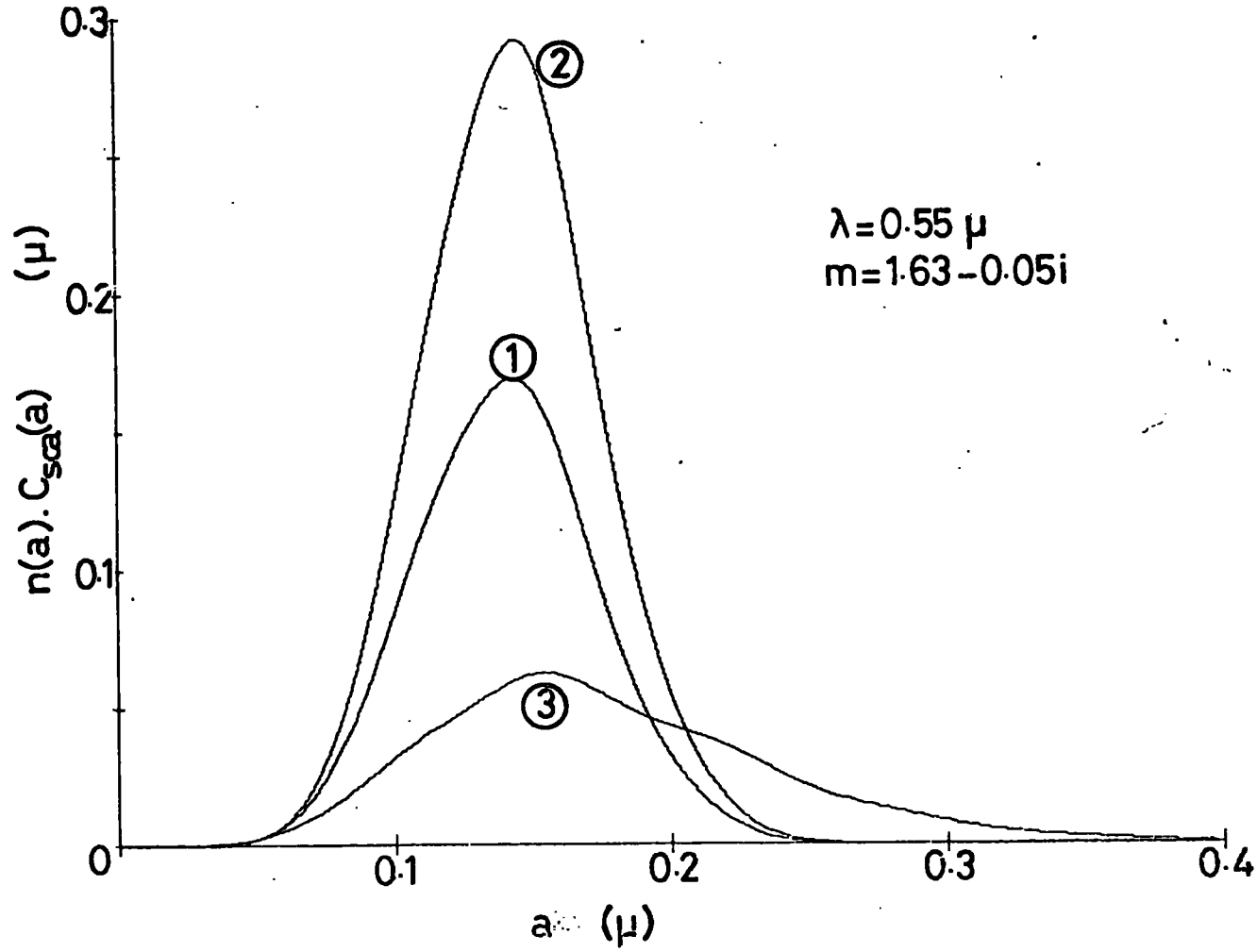


Fig.11·4 Weighted size distributions

of figure 11.4 on the size axis, that the shape of these weighted curves depends on the relative strengths of the increasing grain cross section ($\sim a^2$) and the fall-off of the size distribution for large grain sizes. This fall-off in turn depends on the assumptions about the destruction probability D(a) of equation 11.16. Thus the scattering functions calculated are likely to be mainly affected not by the shape of the size distribution function, but by the functional strength of the fall-off at large grain sizes, and forms using the relatively powerful exponential fall-off are likely to give similar results, for any value of the index within the exponential term within reasonable limits.

With this discussion in mind, the form of equation 11.17 will be used henceforth. This has the advantage of giving a narrow, symmetrical weighted curve, so that the mean scattering size is well-defined, and also allows comparison with much other work which uses this form.

11.9 The Scattering Functions for Various Refractive Indices

The scattering functions have been evaluated, using the size distribution of equation 11.17, for a number of refractive indices which have, at various times, been suggested as typical of the interstellar dust. These are summarised below, the references giving the source of the refractive index for the material quoted:

i) Biological material, m = 1.16 (Hoyle and Wickramasinghe, 1979b)

ii) Ice, m = 1.33 (Wickramasinghe, 1973)

iii) Silicates, m = 1.63 - 0.05i (Gaustad, 1963; Hanner, 1971)

iv) Graphite, m = 2.5 - 1.3i (Wickramasinghe, 1973)

v) Iron, m = 3.55 - 3.0i (Wickramasinghe, 1973)

(A number of recent papers have discussed the possibility of explaining the interstellar extinction law over a wide wavelength range with substances of a biochemical nature, utilising transitions in the π electrons in C - C multiple bonds to explain the observed absorption features. (Wickramasinghe et al., 1977; Hoyle and Wickramasinghe, 1977, 1978, 1979a; Sakata et Al., 1977). Laboratory measurements of biochemical spectra have shown some striking agreements with the measured extinction law. More recently, Hoyle and Wickramasinghe (1979b) have proposed that interstellar grains may be a mixture of bacterial and viral particles and have suggested that the interstellar extinction law can be explained using the laboratory determined mean spectral properties of such particles together with their terrestrial size distributions. Laboratory measurements on freeze-dried bacteria appropriate to the visible waveband give a mean refractive index m = 1.16 (Hoyle and Wickramasinghe, 1979b) and this value has been included in this study

to determine whether it is consistent with the polarisation measurements of reflection nebulae.)

Although, for convenience, grain models based on these refractive indices will be referred to as biological, ice, silicate, graphite and iron grains, it is, of course, impossible to identify a particular grain material on the basis of a single refractive index determination and other evidence must be taken into account before this identification is made. The refractive indices chosen have the advantage of covering a wide range of possible values and it is for this reason, as much as any other, that they are used here.

The resulting scattering functions are shown in figures 11.5 to 11.10, and represent the total scattered intensity:

$$I(\theta) = F_1(\theta) + F_2(\theta) \qquad 11.22$$

and the % polarisation of the scattered light:

$$P(\theta) = \frac{F_1(\theta) - F_2(\theta)}{F_1(\theta) + F_2(\theta)} \times 100 \qquad 11.23$$

The $I(\theta)$ curves are shown only for $m = 1.16$, since the features shown here are similar for all refractive indices considered, while the $P(\theta)$ curves are shown for all 5 refractive indices and for a range of particle

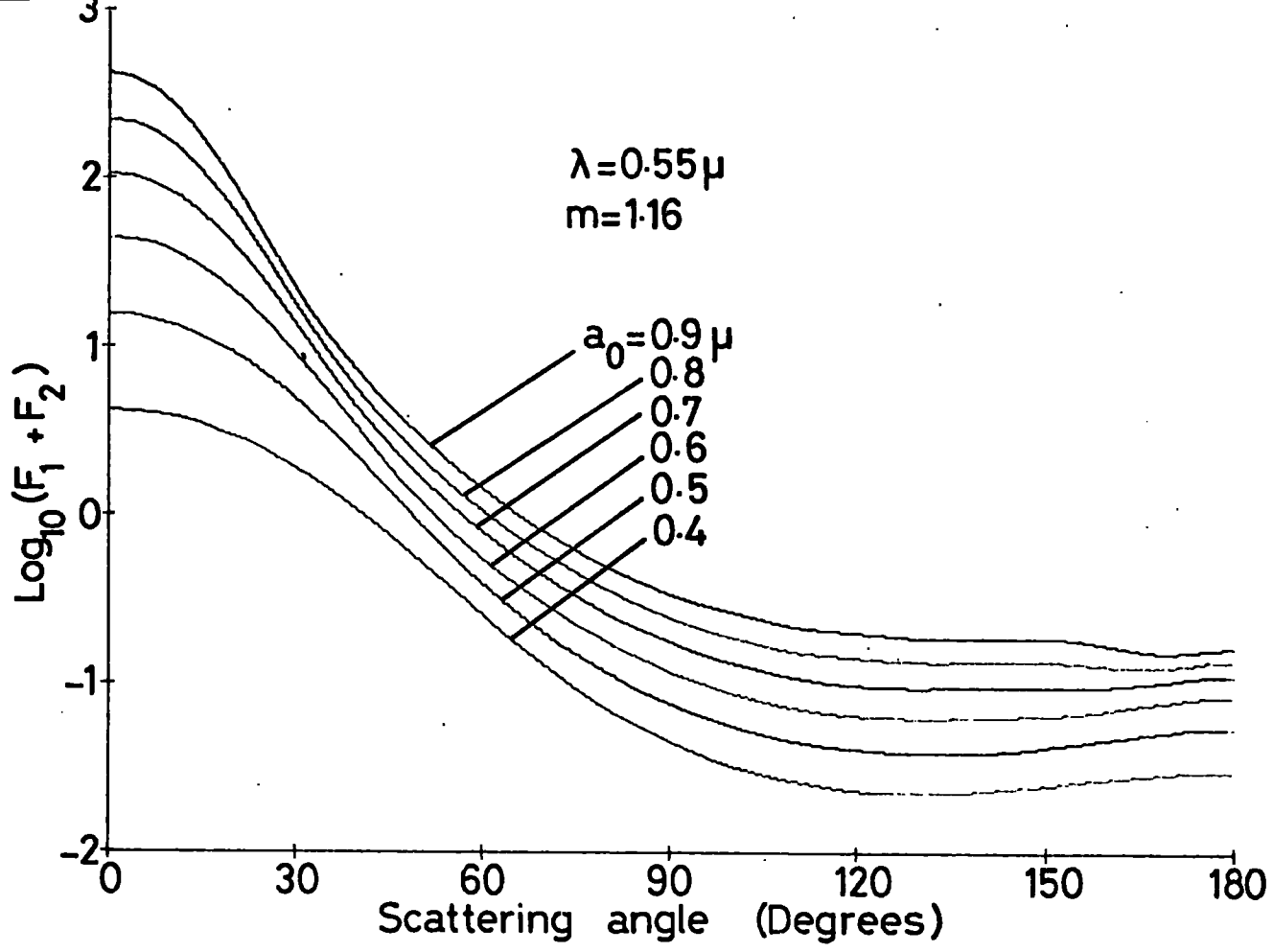


Fig.11.5 Scattered intensity from 'biological' grains

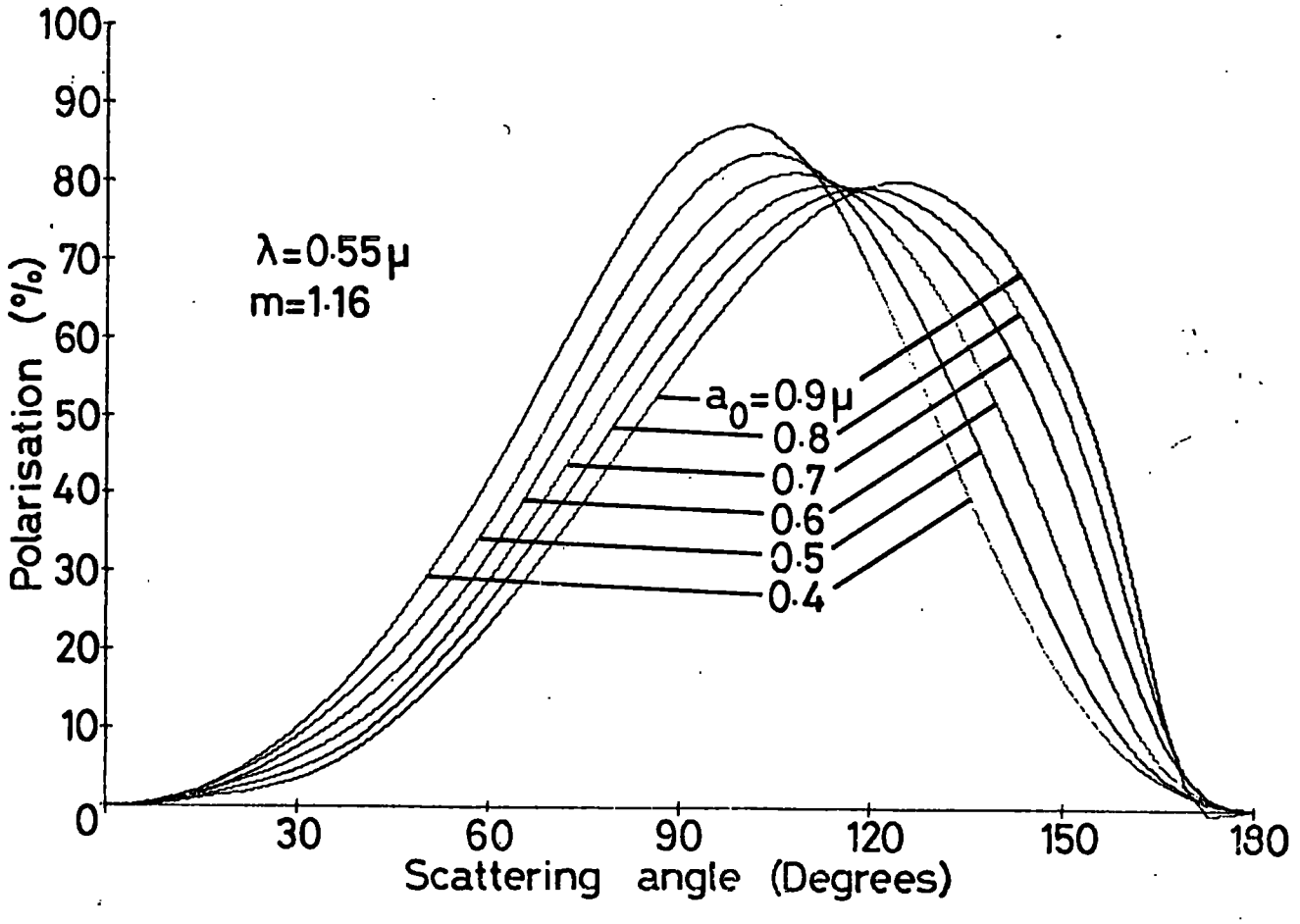


Fig.11.6 Polarisation by scattering from 'biological' grains

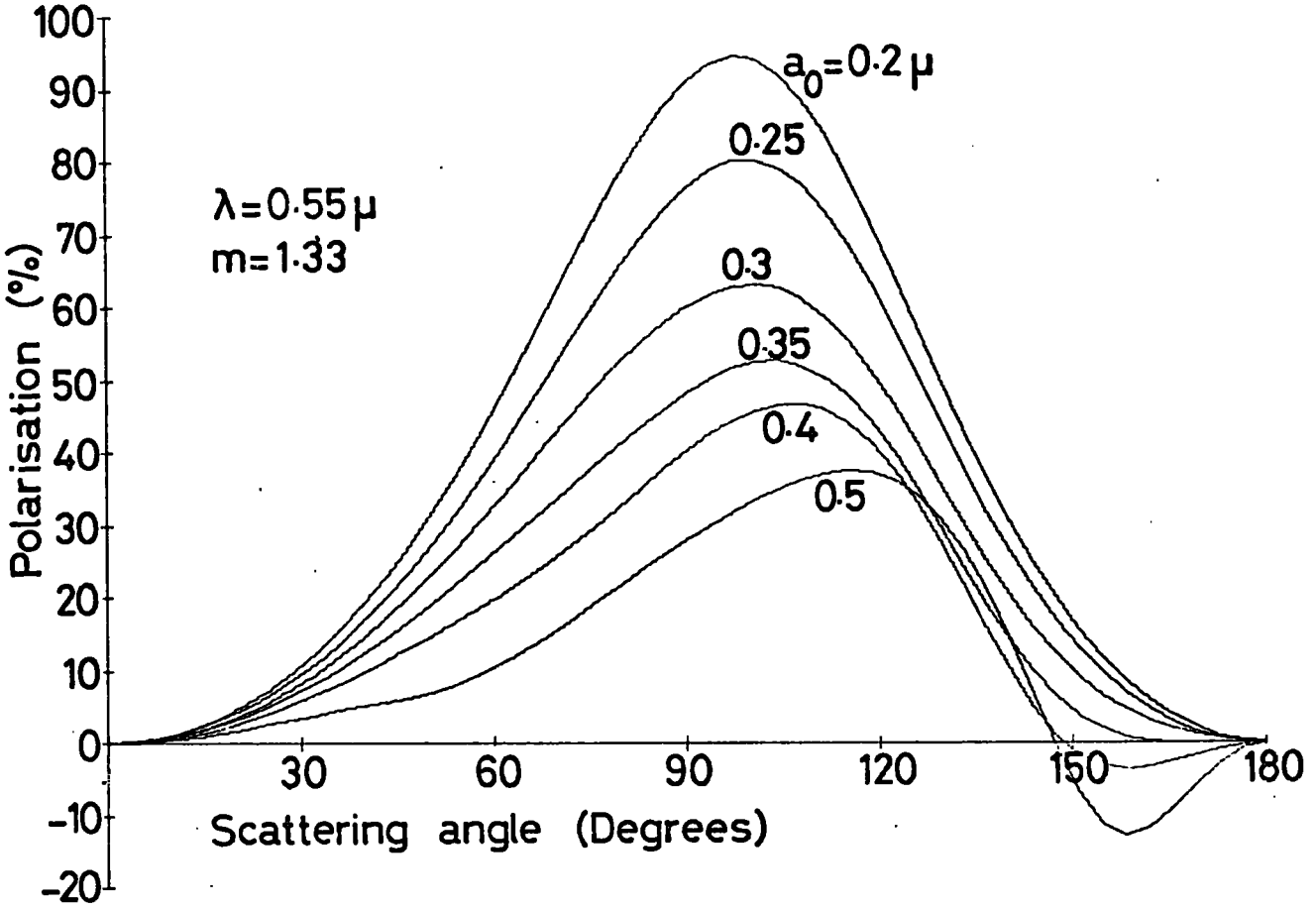


Fig.11·7 Polarisation by scattering from ice

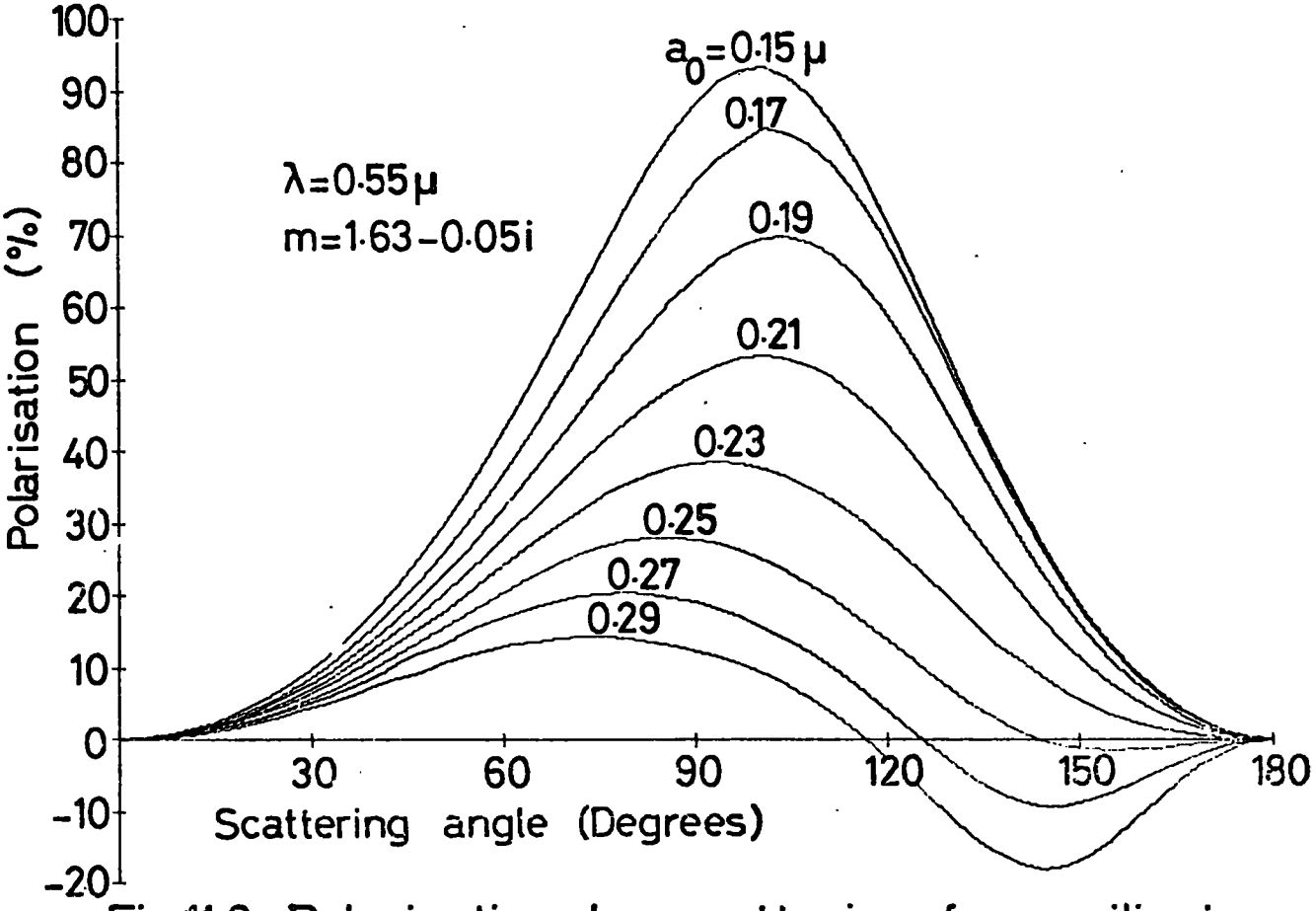


Fig.11·8 Polarisation by scattering from silicates

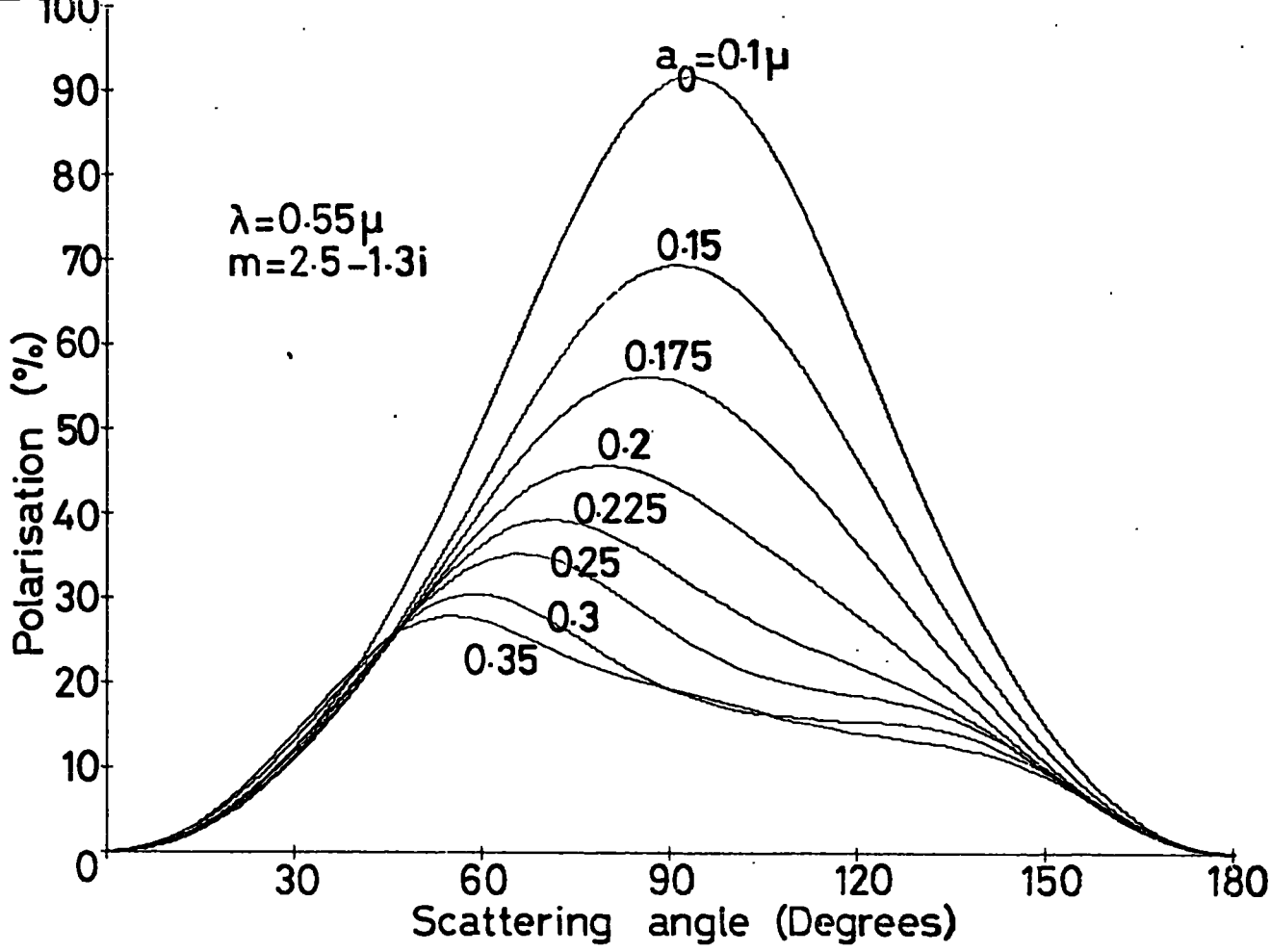


Fig.11·9 Polarisation by scattering from graphite

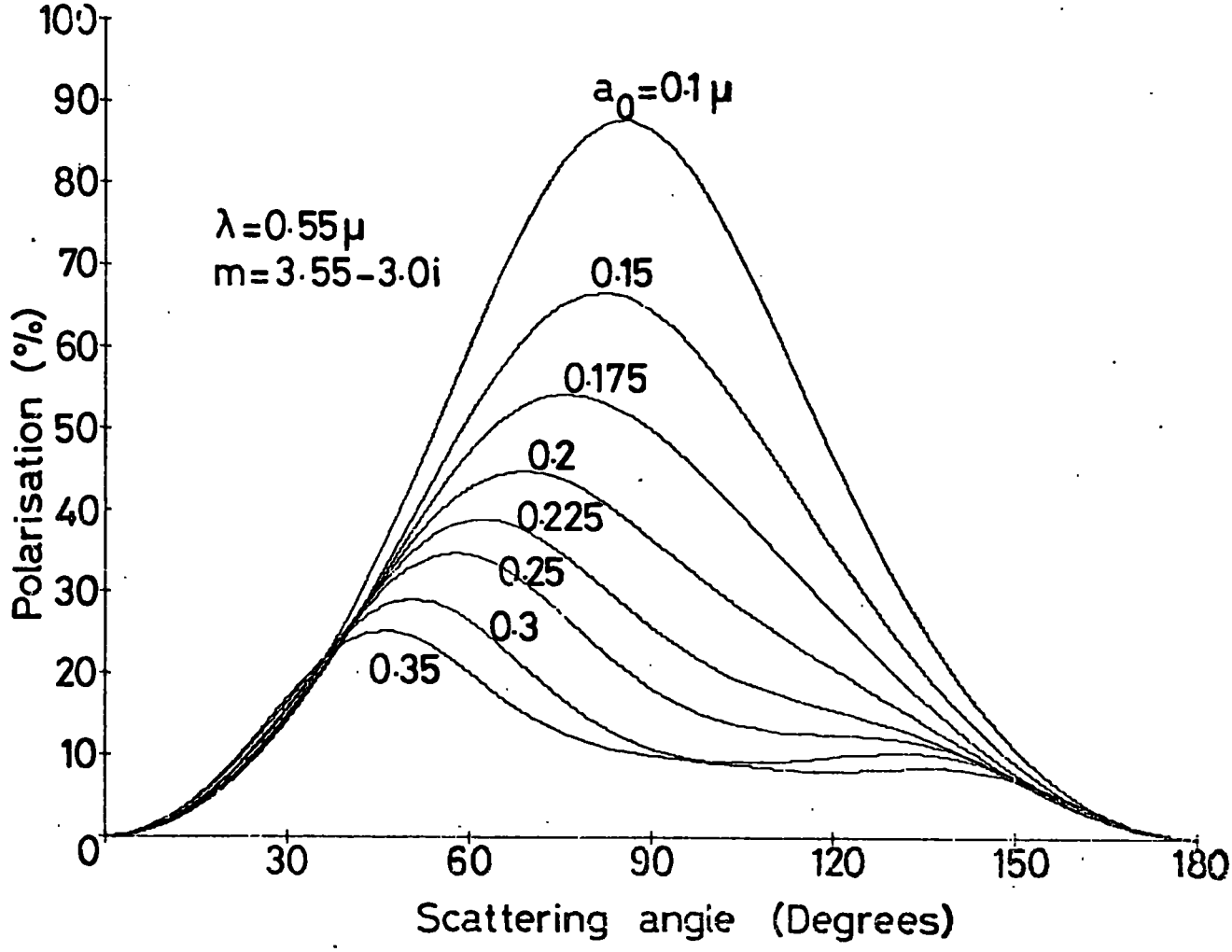


Fig.11·10 Polarisation by scattering from iron

sizes characterised by the parameter a_o of equation 11.17.

A number of general conclusions about the nature of light scattering from sub-micron particles can be drawn from these curves. Most obviously, the scattering functions for a size distribution of particles are far more regular in form than for single spheres and show smooth trends with both mean particle size and scattering angle. The following summarises the main features:

i) On grounds of symmetry, no polarisation is produced for purely forward or backward scattering. Most polarisation is produced for scattering angles close to 90^o.

ii) The scattering functions are all forward - throwing (i.e. most light is scattered through small angles), this being most pronounced for large grain sizes - the scattering becoming more isotropic as the particle size is decreased.

iii) The scattering angle, θ_{MAX}, which gives the largest polarisations, lies above 90^o for small refractive indices and below 90^o for large indices. The asymmetry of the $P(\theta)$ curve increases as the mean grain size increases. For intermediate refractive indices (e.g. silicates), θ_{MAX} lies above 90^o for small sizes and below 90^o for large sizes.

iv) In nearly all cases the polarisation is increased by decreasing the mean grain size (or equivalently, by increasing the wavelength).

v) The average level of polarisation varies very rapidly with particle size, hence polarisation measurements should give an accurate measure of the size of scattering particles in space.

vi) Negative polarisations are produced only with backward scattering (θ near 180°) from large grain sizes. Hence these polarisations could only be observed if forward scattering (which is positively polarised and far more intense) were completely absent.

CHAPTER 12

MODELS OF REFLECTION NEBULAE

12.1 Introduction

Observations of the surface brightness, colour and the wavelength dependence of polarisation of reflection nebulae have been directed, over the years, at a greater understanding of the optical properties of the scattering particles in these nebulae. It is probably true to say, however, that no simple interpretation has emerged which relates the observations unambiguously to the important nebular quantities. More detailed attempts to interpret the observations have relied on the construction of more or less realistic numerical models of the nebula in question and, while such an approach is clearly necessary, the numerical complexity of such models does little to clarify the physical processes at work, so that interpretation of the observations then depends on identifying the important trends in the results of many model calculations.

The difficulty of this approach becomes obvious when one considers the number of parameters which shape the optical appearance of a reflection nebula. These include the (usually unknown) nebular geometry and the refractive index, albedo, size distribution and number density of the scattering dust grains (assuming that the effects of non-sphericity are negligible). In addition, severe numerical problems are involved in calculating the observable quantities when optical depths within the nebula approach

unity and the light from the illuminating star is scattered many times before being observed.

Since the numerical problem of investigating all possible nebular variables is plainly a lengthy process, some method of relating the numerical models to the results of observation is highly desirable so that limits can be set on the range of nebular parameters which need be considered. Due to the complex interaction between these parameters, however, it is not usually a simple matter to set such limits without making assumptions about the remaining unknowns. A methodical approach, linked closely to observations of as many nebular properties as possible, is clearly necessary if these assumptions are to be eliminated.

Until now, the necessary comprehensive data on the brightness, colour and polarisation of both the nebula and illuminating star have been available in only a handful of cases. Perhaps the main reason for this has been that few nebulae are both candidates for accurate observation (i.e. bright) while also being suitable subjects for simple numerical modelling (i.e. optically thin). Most proponents of these nebular studies have advocated further work on the observationally difficult cases of faint, optically thin nebulae, but it is clear that if the theoretical problems of high optical depth can be overcome, the requirement of accurate measurement of as many nebular properties as possible can be more easily

met in the bright nebulae.

In this chapter, theoretical arguments are developed, based on a general optically thin nebular model, which show how polarimetric observations of reflection nebulae at a single wavelength, in conjunction with brightness data, can provide a good indication of both the nebular geometry and some properties of the nebular dust. It is also shown how general arguments, based on simple polarimetric observations, can be used to place limits on the light scattering properties of nebular dust models, which are largely independent of any assumptions about other nebular parameters.

Application of these theoretical arguments to the specific case of NGC1999 is undertaken in chapter 13 and in chapter 14, where the effects of finite optical depth and multiple scattering of the light are investigated. It is found that, unlike surface brightness and colour observations which are very sensitive to the presence of finite optical depth effects, polarimetric observations are largely unaffected, so that deductions based on these, apart from being intuitively more easily understood, are likely to be far more reliable than those based on colour or brightness observations alone.

12.2 Previous Work

Previous theoretical work on reflection nebulae has concentrated both on attempts to elucidate the effects of the main nebular and dust parameters on the observable quantities and on detailed attempts to explain measurements in particular nebulae in terms of specific models. Perhaps the best studied nebula in this respect is the Merope Nebula, for which numerous observations are available (O'Dell, 1965; Elvius and Hall, 1966; Artamonov et Al., 1968).

Roark, Greenberg and Hanner, in a series of four papers (Greenberg and Roark, 1967; Greenberg and Hanner, 1970; Hanner and Greenberg, 1970; Hanner, 1971), have discussed in some detail the colours and polarisation expected for a number of idealised nebular geometries based on plane parallel dust slabs, under the assumption of single scattering (although they allow for extinction effects within the nebula). They used the Mie scattering theory as a basis and derived models to explain the observed colours and polarisations in the Merope Nebula using a variety of dust grain types. They concluded that the observations could best be matched by using ice grains ($m = 1.33$) with $a_o \sim 0.5\mu m$. (Equation 11.17), although the surface brightness data was not explained. Satisfactory fits could not be found using either highly absorbing grains or silicates ($m = 1.63 - 0.05i$).

Studies of idealised models, usually with the assumption of single scattering, have been published by a number of other workers (e.g. Sobolev, 1960; van den Bergh, 1966; Jura, 1975; Rush, 1975; Voshchinnikov, 1977, 1978). Unfortunately, however, in studies not specifically associated with particular nebulae, each worker has tended to make different assumptions (particularly about the scattering properties of the nebular dust). In numerically complex studies such as these, this often makes inter-comparison of results virtually impossible except on the most broad and general features. As a result, the relevance of much of this work to studies of specific nebulae is severely limited. Vanysek (1969) and Zellner (1974) have given reviews of much of this work and also give many more references.

More recently, attempts have been made to include the effects of multiple scattering in nebular models (e.g. Vanysek and Solc, 1973; Witt, 1977a,b,c,d; White, 1979a,b). Further discussion of this work is given in chapter 14.

12.3 A General Nebular Model

Consider the nebula of figure 12.1, illuminated by a star at O with a surrounding nebular medium which scatters light to an observer at a distance D from the nebula. An element of nebular medium of length d_y is shown, lying in a column of cross sectional area A defined

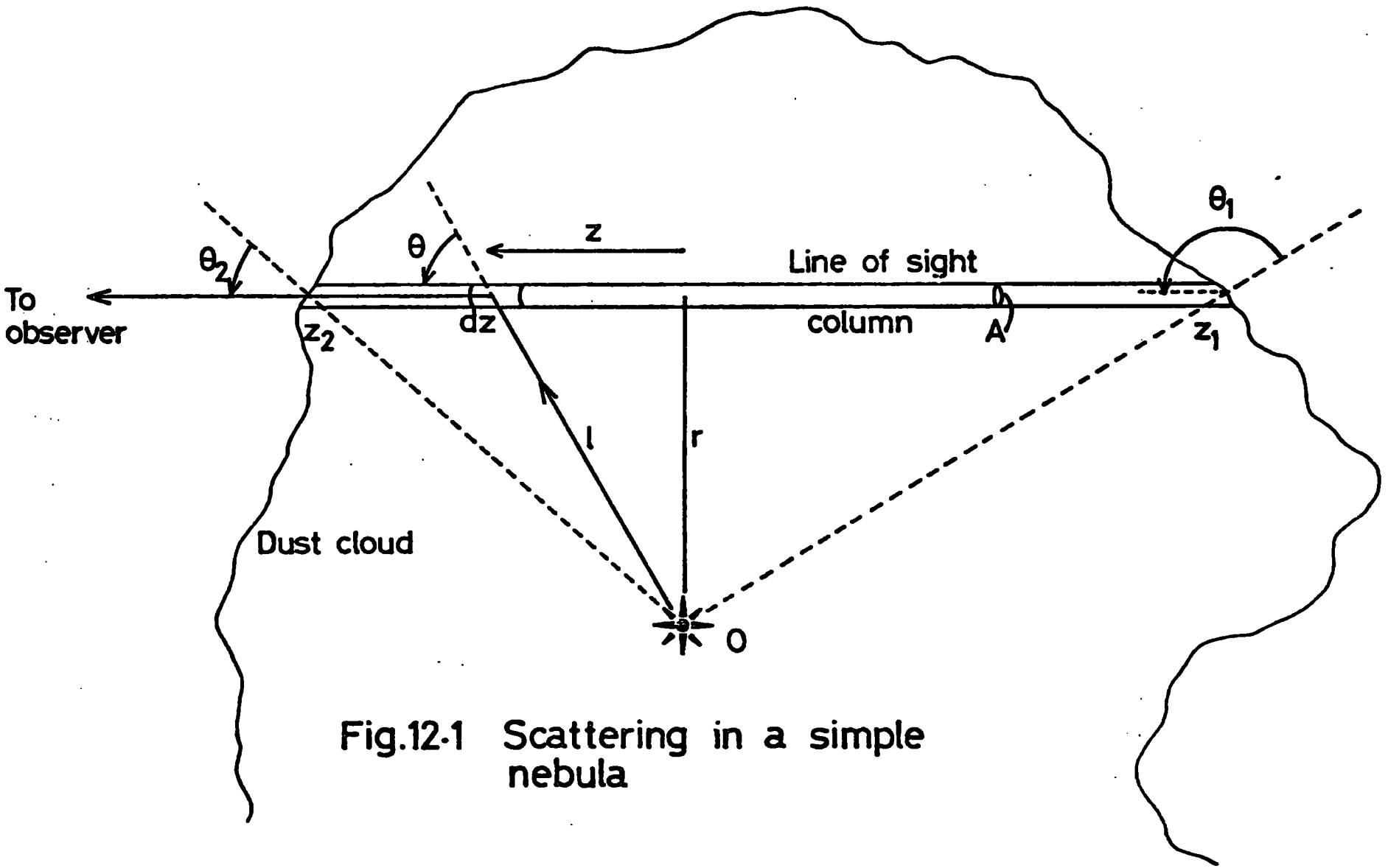


Fig.12-1 Scattering in a simple nebula

by the observer's line of sight through the nebula. At a given wavelength, neglecting extinction, the light flux at this element is:

$$\mathcal{J} = \frac{L}{4\pi \ell^2} \qquad 12.1$$

where L is the stellar luminosity. A single average dust grain in this element scatters a power:

$$p_j = \left(\frac{\lambda}{2\pi}\right)^2 \frac{F_j(\theta)}{2} \mathcal{J} = \frac{L\lambda^2 F_j(\theta)}{32\pi^2 \ell^2} \qquad 12.2$$

into a unit solid angle in the direction of the observer, for each state of polarisation ($j = 1$ for polarisation perpendicular to the scattering plane and $j = 2$ for polarisation parallel to the scattering plane - equations 11.3, 11.12 and 11.15 lead to this result).

If the number density of dust particles is N, the total power scattered by the entire line of sight column is:

$$p_j = \int_{z_1}^{z_2} \frac{L\lambda^2 F_j(\theta)}{32\pi^2 \ell^2} N A dz \qquad 12.3$$

Making a change of variable from z to θ then gives:

$$p_j = \frac{L\lambda^2 A}{32\pi^3 r} \int_{\theta_2}^{\theta_1} N(\theta)\, F_j(\theta)\, d\theta, \qquad 12.4$$

where $N(\theta)$ is the dust density evaluated on the line of sight column at the position where the scattering angle is θ.

To the observer, the area A subtends a solid angle A/D^2 and hence the energy flux at the observer from an element of nebula subtending a solid angle of 1 sq. arc sec. is:

$$\mathcal{F}_o = \frac{p_j}{A\alpha^2} \qquad 12.5$$

where α (= 206,265) is the number of arc seconds in a radian. Hence combining the flux from both states of polarisation the observer sees a total flux:

$$\mathcal{F}_o = \frac{L\lambda^2}{32\pi^3 r \alpha^2} \int_{\theta_2}^{\theta_1} \left[F_1(\theta) + F_2(\theta)\right] N(\theta)\, d\theta \qquad 12.6$$

from each square arc second of nebula. Comparing this with the flux from the star alone (neglecting extinction):

$$\mathcal{F}_* = \frac{L}{4\pi D^2} \qquad 12.7$$

gives the brightness of the nebula, m_{NEB}, in magnitudes per square arc second, in terms of the observed magnitude, m_*, of the star:

$$m_{NEB} = m_* - 2.5 \log_{10} \left\{ \frac{\lambda^2 D^2}{8\pi^2 r \alpha^2} \int_{\theta_2}^{\theta_1} [F_1(\theta)+F_2(\theta)] \, N(\theta) \, d\theta \right\} \qquad 12.8$$

The percentage polarisation of this light is:

$$P = \frac{p_1 - p_2}{p_1 + p_2} \times 100 \qquad 12.9$$

12.4 Conclusions Based on the Optically Thin Model

12.4.1 Nebular Geometry

A number of conclusions can be drawn from the above analysis, together with a knowledge of the form of the light scattering properties of sub-micron particles (Chapter 11), which enable the polarisation and brightness observations of nebulae to be discussed in relation to probable nebular properties. Most important of these are the geometrical considerations.

If the dust density, N, is constant, or (as is likely to be approximately true in many cases) can be described by a power law fall-off with radial distance from the star i.e.:

$$N \propto \frac{1}{l^{\beta}} \quad , \qquad 12.10$$

then the line of sight integral takes the form:

$$\frac{1}{r^{\beta+1}} \int_{\theta_2}^{\theta_1} F_j(\theta) \sin^{\beta}\theta \, d\theta \qquad 12.11$$

and the observed polarisation then depends only on the scattering functions and the limits of integration. Furthermore, since the Mie scattering functions are so forward-throwing (Figure 11.5), it is often mainly θ_2, i.e. the position of the front face of the nebula, which determines the observed polarisation. If this condition holds, it is clear that any variation of polarisation with offset distance from the illuminating star must be explained in terms of the geometry of the nebular boundaries, in particular that of the front face. Three simple geometries exist where no variation of the observed polarisation would be expected, since the angular limits of the line of sight integration are constant, these are: the infinite dust cloud of figure 12.2 (or at least sufficiently large that $\theta_2 \rightarrow 0$ and $\theta_1 \rightarrow 180^{\circ}$), the conical nebulae of figure 12.3 and the special case of a semi-infinite dust cloud with a star on its surface (Figure 12.4).

It is a commonly observed property of reflection nebulae, however, that the % polarisation <u>increases</u> with

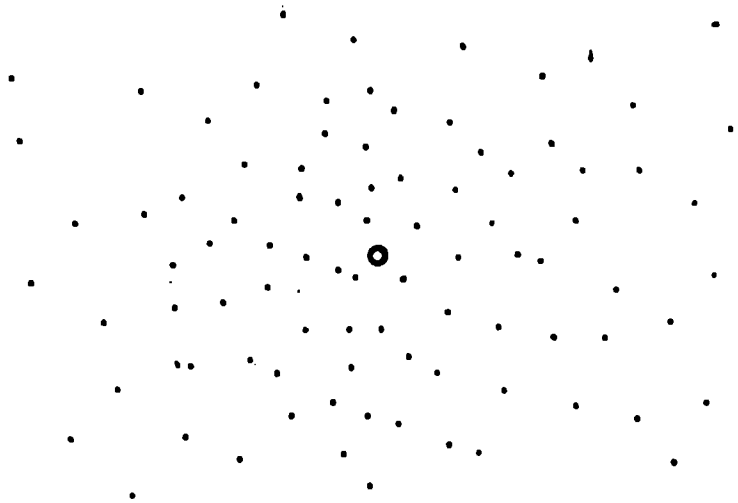


Fig.12.2 Infinite dust cloud

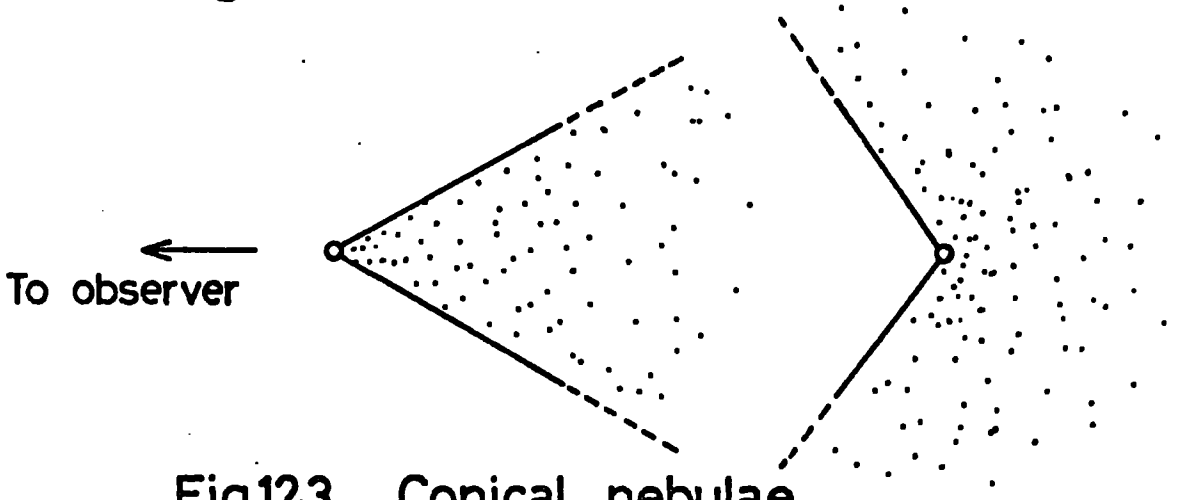


Fig.12.3 Conical nebulae

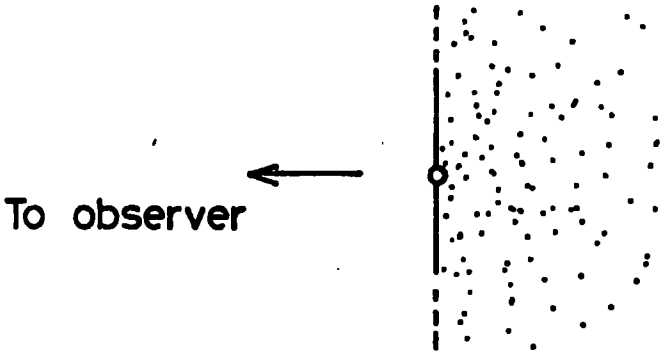


Fig.12.4 Semi-infinite dust cloud

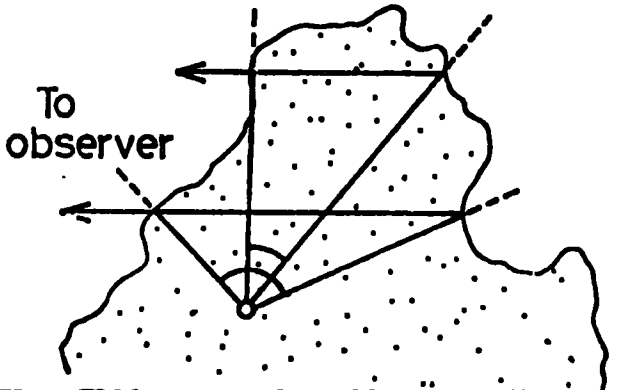


Fig.12.5 Effect of offset distance on the line of sight

increasing offset distance (e.g. NGC 6726/7, figure 9.2; the Merope nebula, Elvius and Hall, 1966; M82, Bingham et Al., 1976; η Carinae, Warren-Smith et Al., 1979). This phenomenon is easily explained by reference to the curves of figures 11.6 to 11.10, which show that the degree of polarisation is enhanced if the range of scattering angles is restricted around 90°. As figure 12.5 shows, a wide range of possible nebular geometries will give a range of integration which becomes more restricted around 90° as the offset distance is increased - indeed it is difficult to devise geometries where this does not occur.

12.4.2 Effects of Dust Density Fall-Off

If the dust density can be approximately described by equation 12.10, then equation 12.11 shows that the effect which this has on the observed polarisation can be explained in terms of giving more weight to the scattering angles around 90°, this weight becoming stronger as β increases, i.e. as the density fall-off becomes faster. Hence, in the presence of a radial dust density fall-off, the observed polarisation should be increased. Conversely, a radial <u>increase</u> in dust density should tend to <u>decrease</u> the observed polarisation, although this situation is unlikely to be common.

The presence of density variations of this form does not, however, alter the conclusion that the <u>variations</u>

in the observed polarisation are caused by the nebular geometry, as discussed in the previous section.

In addition to affecting the polarisation, equation 12.11 shows that the surface brightness fall-off varies as $1/r^{\beta+1}$ (in addition to any geometrical effects due to the changing limits of angular integration) so that observation of the overall surface brightness fall-off may provide information about the presence of dust density variations.

12.4.3 The Observed Range of Polarisation

If the illuminating star is contained within the nebula, then, for small offset distances, the nebula approximates to an infinite dust cloud, since $\theta_1 \rightarrow 180^\circ$ and $\theta_2 \rightarrow 0$. Thus for small offset distances the polarisation expected in an optically thin nebula is obtained by integrating the scattering functions $F_1(\theta)$ and $F_2(\theta)$ over all scattering angles, giving, in general, a non-zero result. As has previously been discussed (12.4.2), radial dust density fall-offs will tend to increase this polarisation. The minimum observed polarisation can thus be used to place a limit on the particle size present, namely: the size parameter a_o cannot be smaller than that which gives the observed minimum polarisation when using a constant dust density and an infinite line of sight. If a_o is smaller than this value, the observed minimum polarisation can only be reproduced by a model

which has the star outside the nebula, or which uses a contrived dust distribution. Figure 12.6 shows how the expected minimum polarisation, P_{MIN}, varies with a_o, where:

$$P_{MIN} = \frac{p_1 - p_2}{p_1 + p_2} \times 100 \qquad 12.12$$

with

$$p_j = \int_0^{180^\circ} F_j(\theta)\, d\theta \qquad 12.13$$

In practice, the minimum polarisation must be measured at a suitable offset distance to avoid contamination by direct starlight - hence the size limit may not be so strict in practice (since the observed "minimum" may be larger than the true minimum), although it is still valid.

An upper limit on particle size can also be obtained since it must be possible to produce the maximum observed polarisation at <u>some</u> scattering angle, i.e. the maxima of the curves in figures 11.6 to 11.10 must be greater than, or equal to, the maximum observed polarisation. Figure 12.7 shows how these maxima vary with a_o.

Taking, for NGC1999, estimates of the minimum and maximum polarisations from figures 10.9 and 10.11 to be 10% and 40% and adjusting the scales of figures 12.6 and 12.7 to a wavelength of 0.5μm., the following limits on a_o result:

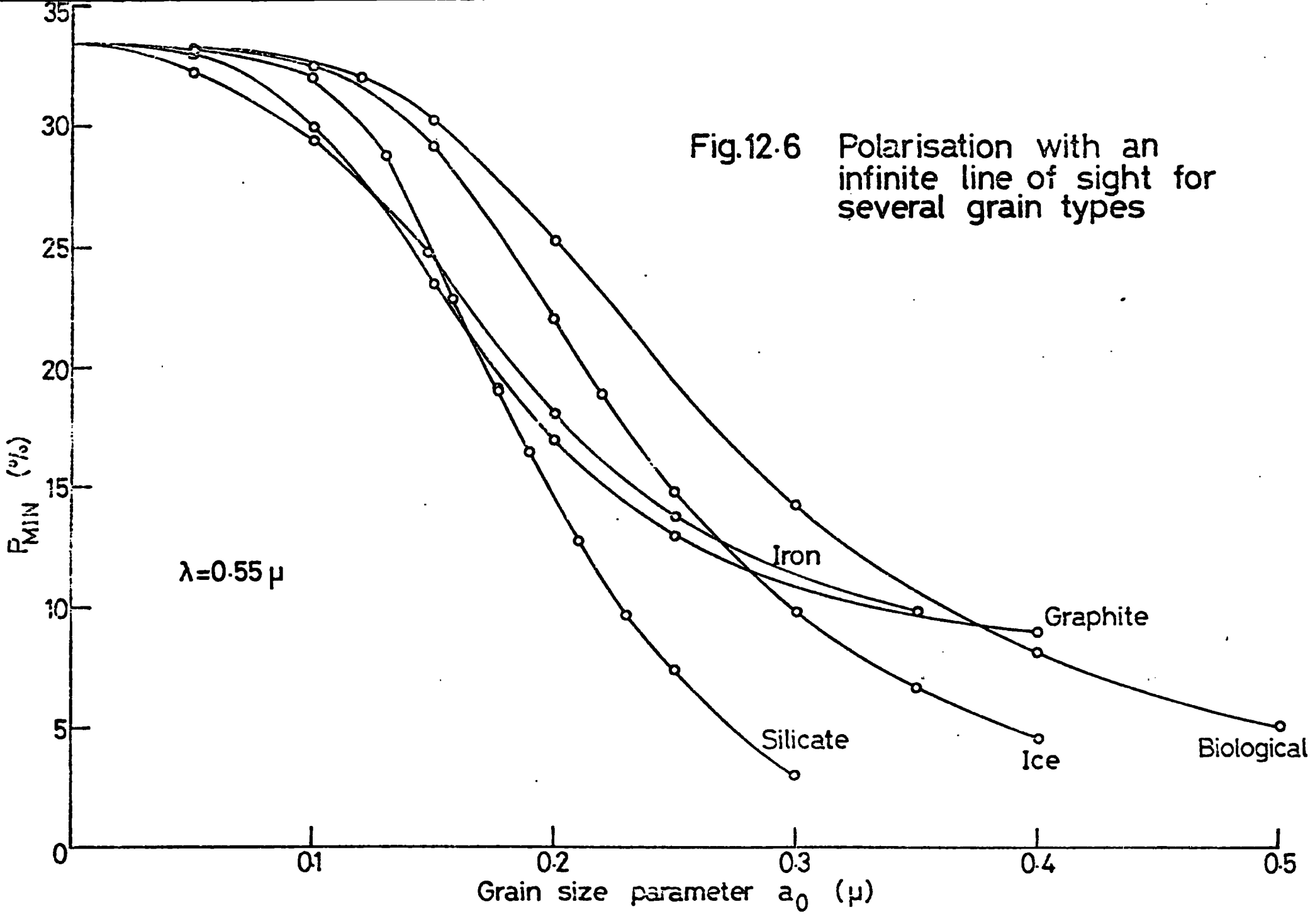


Fig.12·6 Polarisation with an infinite line of sight for several grain types

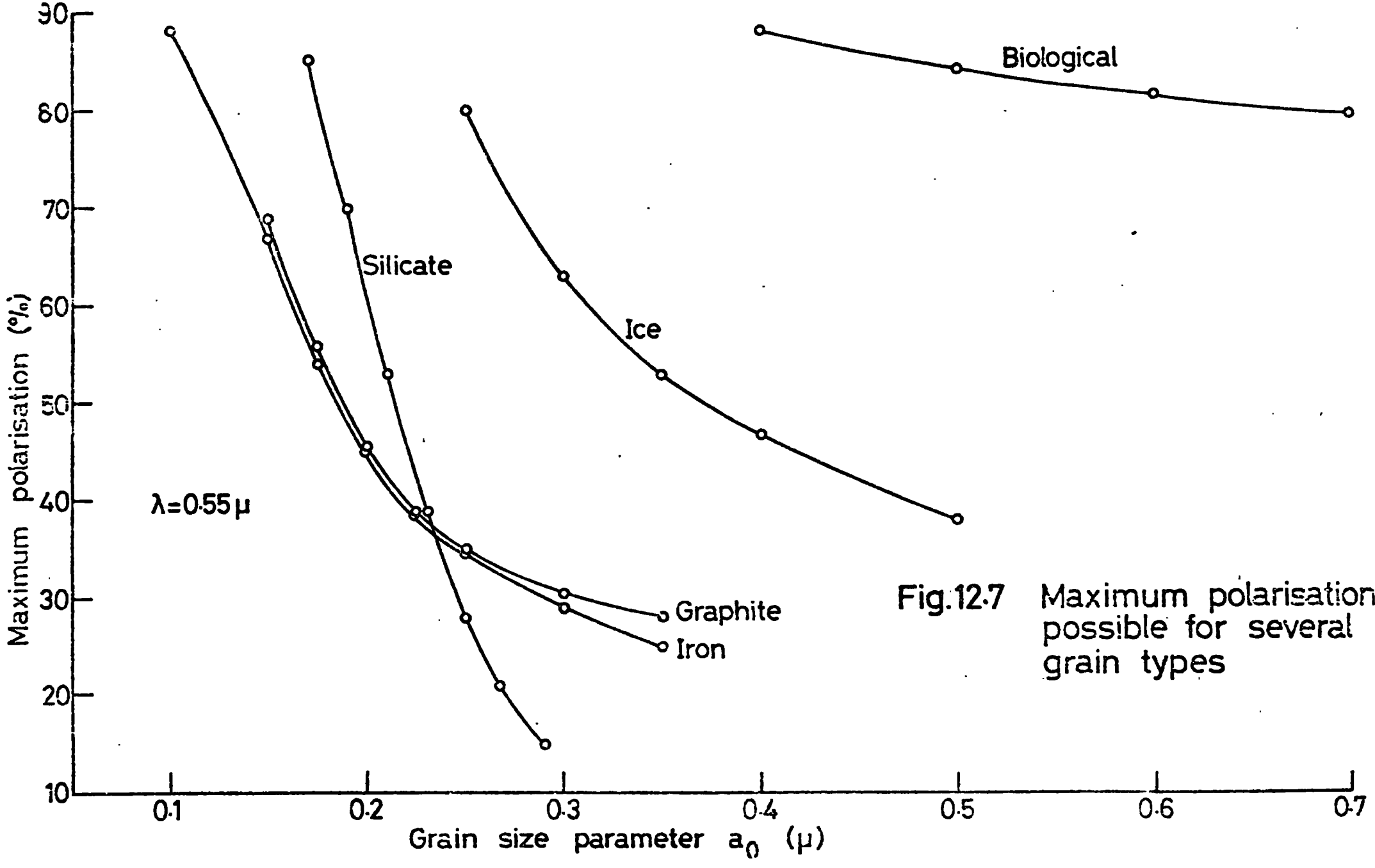


Fig.12.7 Maximum polarisation possible for several grain types

Particle Type	Minimum a_o (μm)	Maximum a_o (μm)
Biological	0.33	VERY LARGE
Ice	0.27	0.42
Silicates	0.21	0.21
Graphite	0.30	0.20
Iron	0.32	0.19

Table 5

On this basis, it is clear that iron and graphite grains cannot explain the observations, while silicate grains can do so only marginally.

While the upper limit on a_o remains valid even in the presence of finite optical depths, it is likely that depolarisation by multiple scattering may mean that the lower limit is inappropriate in some cases. A more general criterion based on this principle will be formulated when the effects of finite optical depth have been discussed (12.6).

12.5 Finite Optical Depth Effects

12.5.1 Internal Extinction

The effects of finite optical depth in a nebula may be considered in two ways - approximately, in terms of light lost from "rays" propagating within the nebula, i.e. internal extinction, or by using a full description

of the multiple scattering of light which occurs in a dense scattering medium.

Extinction operates by two processes: absorption of light energy resulting in heating of the dust grains and scattering of light energy out of the beam. If the number density of scattering centres is N, a beam of light is extinguished such that its intensity after propagating a distance t is reduced from i_o to:

$$i = i_o \exp(-t/t_o) \qquad 12.14$$

where the extinction length, t_o is given by:

$$t_o = (NG_{EXT})^{-1} \qquad 12.15$$

If, at any point in the nebula, t_o is comparable to, or smaller than, typical nebular dimensions, then the effects of finite optical depth must be considered. The effects of extinction are easily included in the formulae of section 12.3 by including a factor to allow for the extinction of light on its passage from star, to scattering centre, to observer (Figure 12.1). Equation 12.4 then becomes:

$$p_j = \frac{L\lambda^2 A}{32\pi^3 r} \int_{\theta_2}^{\theta_1} N(\theta)\, F_j(\theta)\, \gamma(\theta)\, d\theta \qquad 12.16$$

where the extinction factor, $\gamma(\theta)$, is given by:

$$\gamma(\theta) = \exp\left\{-\int NG_{EXT}\, dt\right\} , \qquad 12.17$$

the integral being performed over the light path corresponding to the scattering angle θ. If the dust density is constant, this has the form:

$$\gamma(\theta) = \exp\left\{-NG_{EXT}\, r\, (\sec\theta - \cot\theta + \cot\theta_2)\right\} \qquad 12.18$$

From the above equations it follows that, for all spherically symmetric dust density distributions, less extinction occurs for light scattered close to the front face of the nebula than for that scattered further in. As a result, the importance of the front face in determining the polarisation pattern observed is enhanced by the presence of finite optical depths - in fact, if the optical depth is sufficiently high, the back face may be completely unimportant. Since increased weight is given to the forward scattering angles, the presence of extinction will tend to decrease the observed polarisation if forward scattering angles are present (See figures 11.6 to 11.10), but to increase the observed polarisation if only backward scattering is important.

The observed surface brightness gradient is expected

to become somewhat steeper in the presence of extinction and, since the extinction length is usually longer for red light than for blue, the fall-off will be more rapid for blue light, resulting in a progressively redder nebula as the offset distance is increased. The effects of multiple scattering of the light are, however, likely to make an important contribution to this effect.

12.5.2 Multiple Scattering

A simple theoretical discussion of the effects of multiple scattering is not possible at this stage. A detailed numerical investigation will be found in chapter 14, however, some simple conclusions can be made here.

Multiple scattering becomes important when a non-negligible fraction of the light is scattered within the nebula and then becomes available to be re-scattered before leaving the nebula. Thus, multiple scattering is important when the dust density is high and the grains have a high albedo (i.e. they are not very absorbing). Both of these features result in a high surface brightness, which may thus be a good indication that multiple scattering is important. Some of the effects expected in such cases are:

i) Since the multiply scattered light will not be propagating radially from the star, the polarisations produced by light scattered to the observer from different directions will tend to cancel, producing overall lower

polarisations. This effect is not as large as many workers have suggested, however, since the strongly forward throwing scattering functions (e.g. figure 11.5) result in a radiation field which is still approximately radial, even after several scatterings. A large number of scatterings are required before the radiation becomes isotropic.

ii) Because multiple scattering involves the scattering of light which is already polarised, circular and elliptical polarisations can be produced. To make this observable, however, some degree of geometric asymmetry in the nebula would be necessary. The amounts of circular polarisation expected are, in any case, small.

iii) If the grain albedo is high, most of the light eventually leaves the nebula without being absorbed. Hence the conclusions concerning surface brightness and colour in the presence of extinction (which assumes that all the light is lost on its first scattering) are upset by the effects of multiple scattering. In general the nebular brightness will be higher and the colour variations less pronounced.

iv) Because the light which is seen at large offset distances from the star has undergone the largest number of scatterings, the effects of multiple scattering are expected to be most noticable at large offset distances.

12.6 The Polarising Range of Dust Grains

It was noted in section 12.4.3 that iron and graphite grains could not explain the observed range of polarisation (10% to 40%) in NGC1999. However, in view of the probable effects of multiple scattering in causing depolarisation of the light observed, it is evident that the limits of table 5 may not be valid if finite optical depths are present. To formulate a more general criterion, therefore, define the "polarising range" of a given grain type and size distribution by:

$$\eta = \frac{P_{MAX}}{P_{MIN}} \qquad 12.19$$

where P_{MAX} and P_{MIN} are the polarisations of figures 12.7 and 12.6.

While this ratio does not set an absolute limit on the ratio of maximum to minimum polarisation which can be observed, it is unlikely that this ratio will be exceeded, if the star is contained within the nebula, for the following reasons:

i) The maximum observed polarisation cannot exceed P_{MAX}, and will in most cases be considerably less, due to the effects of a finite line of sight and the presence of multiple scattering.

ii) If the star is contained in the nebula so that the line of sight integration includes the range of scattering

angles close to 90^{o}, the observed minimum polarisation can only be made greater than P_{MIN} by including the effects of a non-infinite line of sight or the presence of spherically symmetric dust fall-offs.

iii) The only simple effect which can decrease the minimum observed polarisation below P_{MIN} - multiple scattering - is expected to have a proportionately larger effect in decreasing the observed maximum polarisations (which are seen at larger offset distances), so long as variations in the amount of multiple scattering from point to point (i.e. variations in dust density) are not too large. This expectation is born out by the detailed multiple scattering calculations of chapter 14 and also by the work of White (1979a,b).

As a result, any dust grain model which predicts a value of η less than the observed ratio of maximum to minimum polarisations, is unlikely to form the basis of a successful model unless special effects are included to enhance this ratio. Figure 12.8 shows the calculated values of η for the 5 grain materials considered previously, together with the upper size limits of table 5 (appropriately adjusted to a wavelength of 0.55μm), which remain valid even in the presence of multiple scattering.

Physically, the parameter η describes the amount of depolarisation produced by averaging the scattering functions over all scattering angles. Thus grains with very

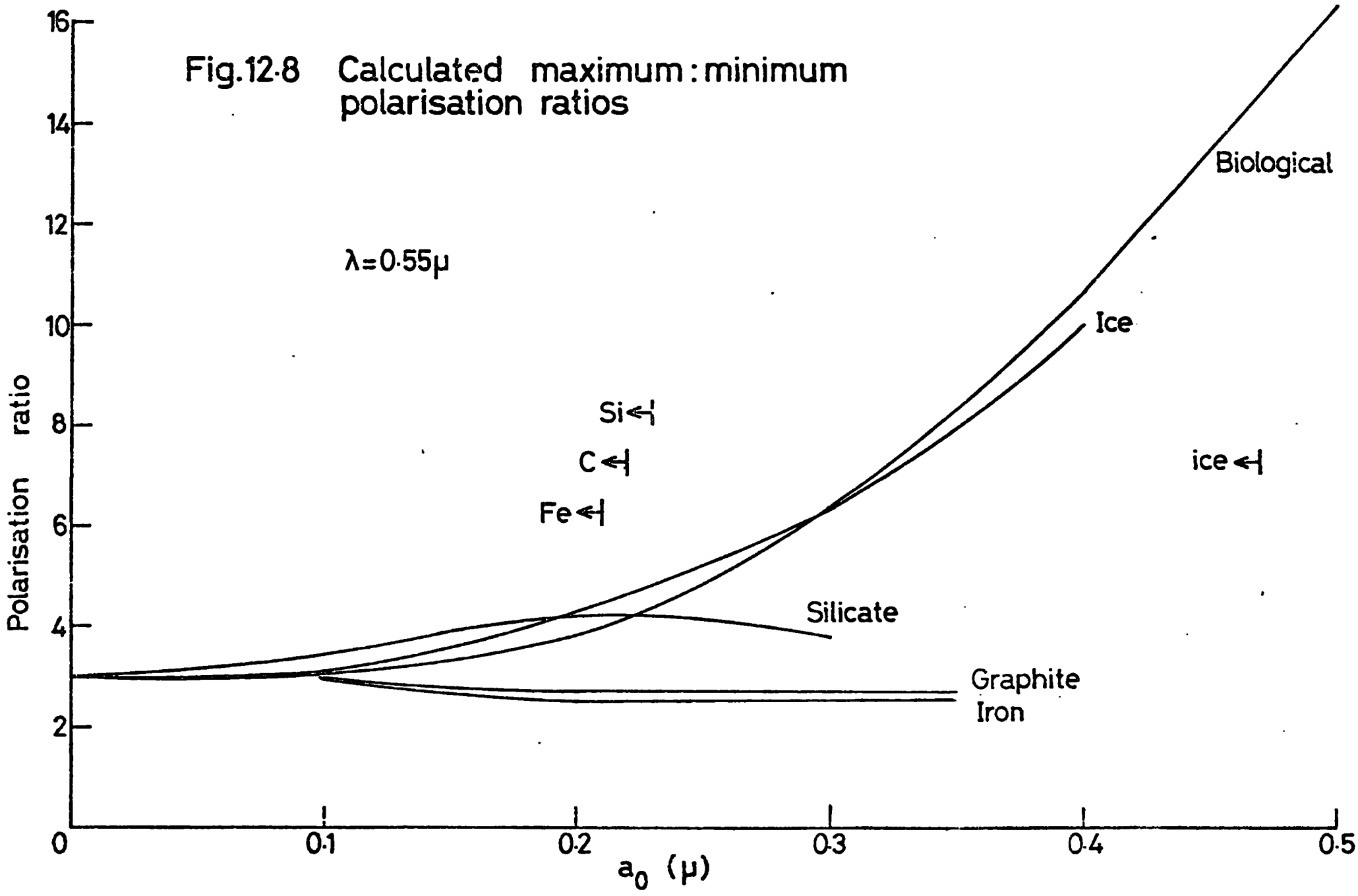


Fig.12·8 Calculated maximum : minimum polarisation ratios
λ=0·55μ
Biological
Ice
Si
C
Fe
ice
Silicate
Graphite
Iron
Polarisation ratio
16
14
12
10
8
6
4
2
0
0·1
0·2
0·3
0·4
0·5
a_0 (μ)

forward-throwing scattering functions give large values of η since most of the light is scattered through small angles where little polarisation is produced. Conversely, grains with isotropic scattering functions give small values of η. As such, this parameter may provide an observational basis for rejecting certain grain models.

12.7 Summary of Conclusions

Following the discussions in this chapter, the following table summarises the most important relationships between the nebular parameters and observable quantities:

Observable Quantity	Strongly Dependent on:	Relatively Weakly Dependent on:
Mean level of polarisation, %	Grain size and dust density gradients	Dust irregularity, overall dust density, small optical thicknesses
Spatial structure of % polarisation	Geometry, particularly of the front face of the nebula	Grain size, optical thickness, dust density variations
Overall range of % polarisation	Grain type and size, dust density gradients	Geometry, small optical thicknesses
Surface brightness	Grain albedo, dust density	Grain size, geometry
Surface brightness gradient	Dust density gradients, internal extinction, degree of multiple scattering	Grain size, geometry

Table 6

CHAPTER 13

A MODEL OF NGC1999

13.1 Introduction

The observations of NGC1999 (Chapter 10) suggest that a model reflection nebula based on a simple geometry and incorporating an appropriate tilt with respect to the plane of the sky may be able to explain the polarimetric observations. In this chapter, an initial model of this sort is developed, based on the conclusions of chapter 12 relevent to optically thin nebulae, and trial calculations are performed to determine the range of geometric parameters and the dust size and type appropriate to explain the observations.

In the later sections of this chapter, this initial model is refined to include the effects of extinction of light within the nebula and the use of a relatively broad optical bandwidth for the observations and is interpreted in the light of its ability to also explain the surface brightness observations.

Further refinement of the model, to include the effects of multiple scattering of the light, is discussed in chapter 14 and these calculations lead to estimates of the nebular colour (Chapter 15) which are then compared with the measurements of Brück (1974). Hence the final model attempts to simultaneously explain the polarisation, surface brightness and colour measurements currently available.

13.2 Relating the Observations to Possible Models

The observations of chapter 10, in conjunction with the conclusions reached in chapter 12 concerning the parameters of nebular models relevant to the polarisation observations, suggest that the following features may be important in a potential nebular model:

i) In order to explain the radial increase in % polarisation, the nebular geometry, in particular the front face of the nebula, must be such as to restrict the range of scattering angles around (or backwards of) 90^{o} as the offset distance is increased.

ii) The higher polarisations in the NE quadrant are most naturally explained in terms of a geometric asymmetry which provides additional restriction of the scattering angles around (or backwards of) 90^{o} in one quadrant. This may be due to a tilt with respect to the plane of the sky, but such an effect cannot be so large as to introduce much asymmetry into the surface brightness's fall-off.

iii) Since the overall trends in the polarisation data are smooth, the nebular geometry must be represented by smooth shapes.

iv) To explain the overall level and range of polarisation observed, the size of the dust grains must be of the order shown in table 5, although the depolarising effects of multiple scattering may mean that these

sizes must be reduced slightly.

v) The observed ratio of maximum to minimum polarisation of ~ 4 suggests that small refractive index particles with a $P(\theta)$ curve which peaks at backward scattering angles may be most successful in explaining the observations. (From a comparison of figure 12.8 with figures 11.6 to 11.10).

vi) The high surface brightness suggests that the effects of multiple scattering may prove important.

13.3 Possible Geometries

A number of simple geometries initially appear to satisfy the above criteria and so will be discussed in more detail.

The spherical nebula of figure 13.1, where an asymmetry is introduced by displacing the star from the centre of the sphere, is one possibility. The sphere, however, must be sufficiently large to explain the visual extent of the nebula and with such a size very little asymmetry is introduced at small offset distances. Furthermore, the geometric effect which a sphere has on the surface brightness is quite large, since the length of the line of sight decreases rapidly towards the edge, while the long line of sight in front of the star may introduce appreciable stellar reddening, which is not observed. Such a nebula must also be considered distinct from the

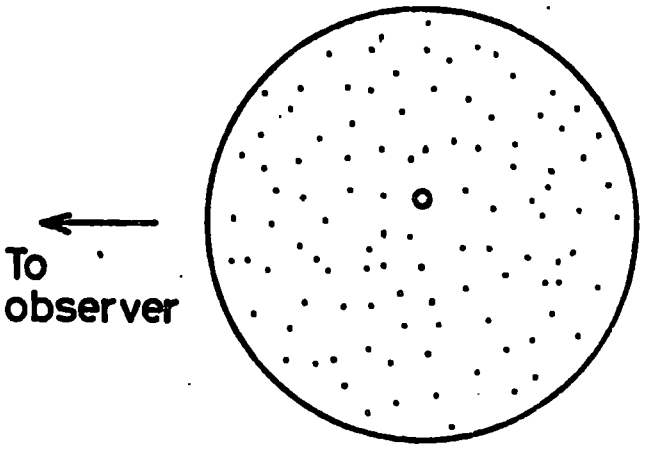


Fig.13·1 Spherical model

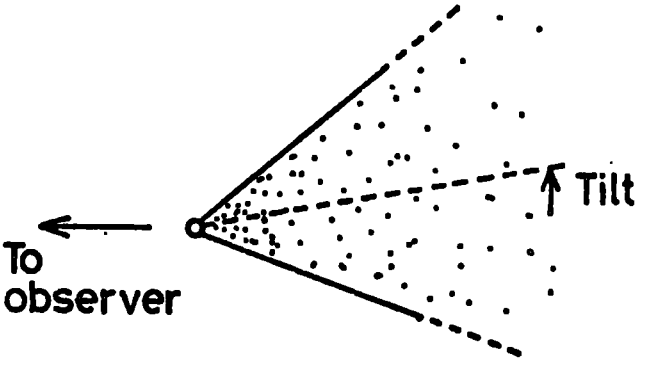


Fig.13·2 Head-on cometary model

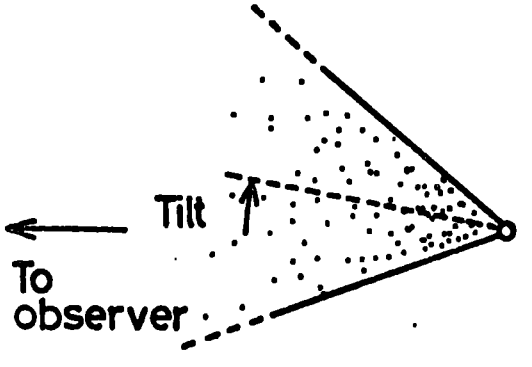


Fig.13·3 Tail-on cometary model

Orion cloud, while the observations of Herbig (10.2.2) naturally suggest an association.

Brück (1974), arguing that NGC1999 shows similar colours to NGC2261, a cometary nebula associated with the star R Mon., has suggested that NGC1999 may be a cometary nebula, seen head-on. In view of the polarisation data, a tilt must presumably also be included (Figure 13.2). Nebular colours, however, are likely to be better indicators of dust size and optical depth effects than of geometry, and the greater sensitivity of polarisation measurements to geometric factors means that this model may be considered inadequate on a number of grounds. Most importantly, there is little or no change in the range of scattering angles as the offset distance is increased, so that the observed large range of polarisation variation cannot be reproduced. Furthermore, only very small polarisations can be produced using typical interstellar-sized dust particles unless the cone angle is made untypically large for cometary nebulae. Finally the predicted surface brightness is likely to be low since the backward scattering involved is highly inefficient and the nebula also subtends a small solid angle at the star.

Modification of this model, a cometary nebula seen tail-on, (Figure 13.3) may go part of the way to solving the surface brightness problem (although the solid angle at the star is still small) but introduces the problem of excessive stellar reddening. Both cometary models also

require the nebula to be distinct from the Orion dark cloud.

The most promising type of nebular geometry is one based on large dust slabs. Such geometries arise naturally whenever a dust cloud (e.g. the Orion cloud) has spatial structure appreciably larger than the visible nebulae contained within it, so that the cloud boundaries can be approximated locally by planes. Such an approach has been used with some success by Greenberg and Roark (1967), Greenberg and Hanner (1970), Hanner and Greenberg (1970) and Hanner (1971) to explain the observations of the Merope nebula in terms of a dust slab 1.0 pc. thick, inclined at 10^{o} to the plane of the sky and with a star 0.35 pc. beneath its surface. The semi-infinite dust slab model of figure 13.4 could arise naturally in the case of NGC1999 through having an illuminating star embedded below the surface of a very large dust cloud (consistent with the observed lack of background stars in this region) and this model also has a number of attractive features as regards explaining the observational data. Polarisations of the size observed can be produced using dust grains of typical interstellar size and the reduction in forward scattering angles at increasing offset distances naturally explains the observed rise in polarisation. The required asymmetry, arising as a result of a tilt of the front surface of the dust cloud, gives rise to a polarisation which is asymptotically

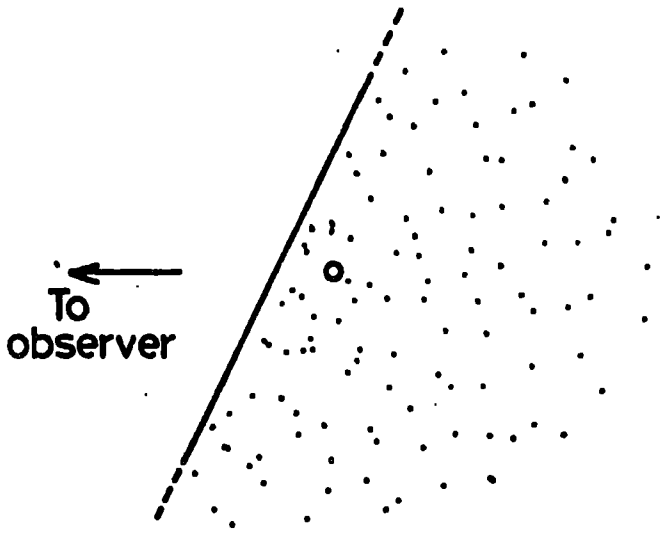


Fig.13.4 Dust slab model

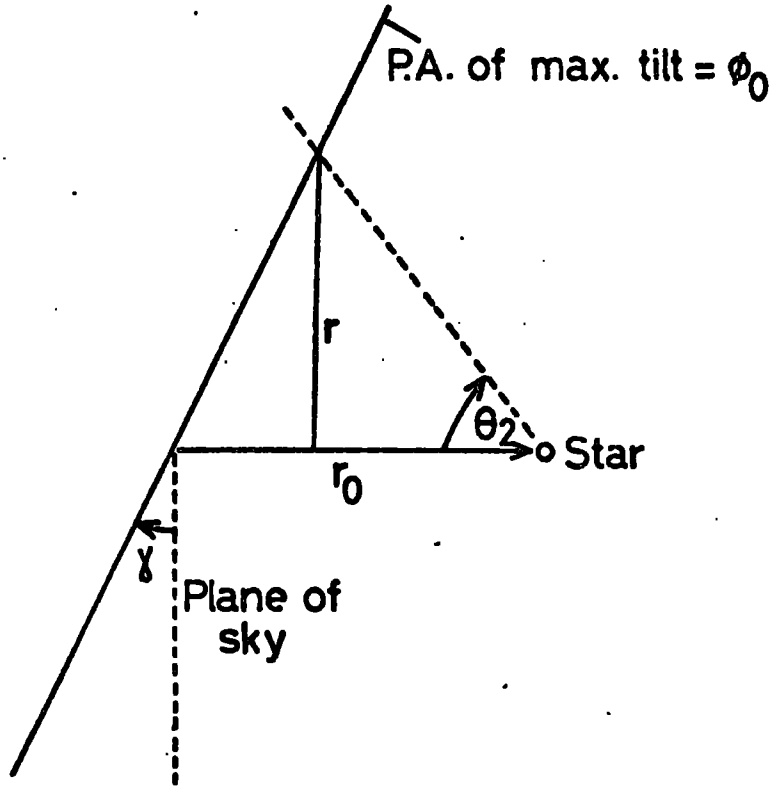


Fig.13.5 Geometrical details of the model

different in each direction, as observed, while the length of the line of sight is everywhere large leading to small geometric effects on the surface brightness. Furthermore, the nebula subtends a large solid angle at the star but does not introduce excessive reddening and the nebula need no longer be considered as distinct from the Orion cloud.

13.4 Details of the Preliminary Model

13.4.1 Geometric Details

Since this latter geometry appears to be the most likely to explain the observed features of NGC1999, a numerical model was constructed to test how satisfactory the explanation was and to determine the parameters required of the geometry and the dust grains.

Figure 13.5 shows details of the geometrical formulation of the model for this purpose. The geometry is characterised by 3 parameters: r_o - the depth of the star behind the front face of the nebula, γ - the tilt of the front face of the nebula with respect to the plane of the sky and ϕ_o - the position angle at which the tilt occurs. Figure 13.5 shows that the angular limits for the line of sight integration at an offset distance r are:

$$\theta_1 = \pi \qquad 13.1$$

$$\theta_2 = \arctan \left\{ \frac{r}{r_o - r \tan \gamma \cos (\phi - \phi_o)} \right\} \qquad 13.2$$

where ϕ is the position angle relative to the star.

Since this represents only a preliminary investigation, the simplifying assumptions of constant dust density, negligible extinction within the nebula and a single wavelength of 0.5μm have been made. With these assumptions, the model polarisation at any point can be found from:

$$P = \frac{p_1 - p_2}{p_1 + p_2} \times 100 \qquad 13.3$$

with:

$$p_j = \int_{\theta_2}^{\theta_1} F_j(\theta)\, d\theta \qquad 13.4$$

13.4.2 Numerical Details

To obtain a set of polarisation data uncorrupted by errors, to which the model can be fitted, a polarisation map was produced using integration areas 1.7" square (5 x 5 pixels) and those measurements with an error on orientation greater than 30°, a deviation from the circular pattern greater than 3 x the standard deviation expected, a total intensity below 5 PDS units (~22.5 mag. per sq. arc sec.) or a distance from the star of less than 8" were rejected. In addition, two regions, one centred on the dark obscuration and the other on a region due south of V380, where small polarisations are

measured, were omitted. This latter region does not appear to form part of the overall pattern and as such the model does not represent it. However, such deviations are to be expected in any real nebula.

Figure 13.6 shows the measurements used. The number rejected is small and the overall features to be explained are still apparent. The 4 model parameters: r_o, γ, ϕ_o and a_o, the grain size parameter, have been adjusted to give the best fit to the data by minimising:

$$\chi^2 = \frac{1}{n} \sum \frac{(P_m - P_{OBS})^2}{\left[\text{Max}(\sigma,\ 5\%)\right]^2} \qquad 13.5$$

where P_m and P_{OBS} are the model and observed polarisations and n is the number of data points. The weights used in the fitting process were derived from the measurement error, σ, but with a lower limit of 5% applied. In practice nearly all the data points are given equal weight by this technique.

In performing the minimisation, it was found that the grain size parameter, a_o, was well-defined by the data and the geometrical parameters could not be adjusted to obtain a fit whatever the value of a_o. The integrations of equation 13.4 were performed using the trapezium rule, with a step length of 1^o.

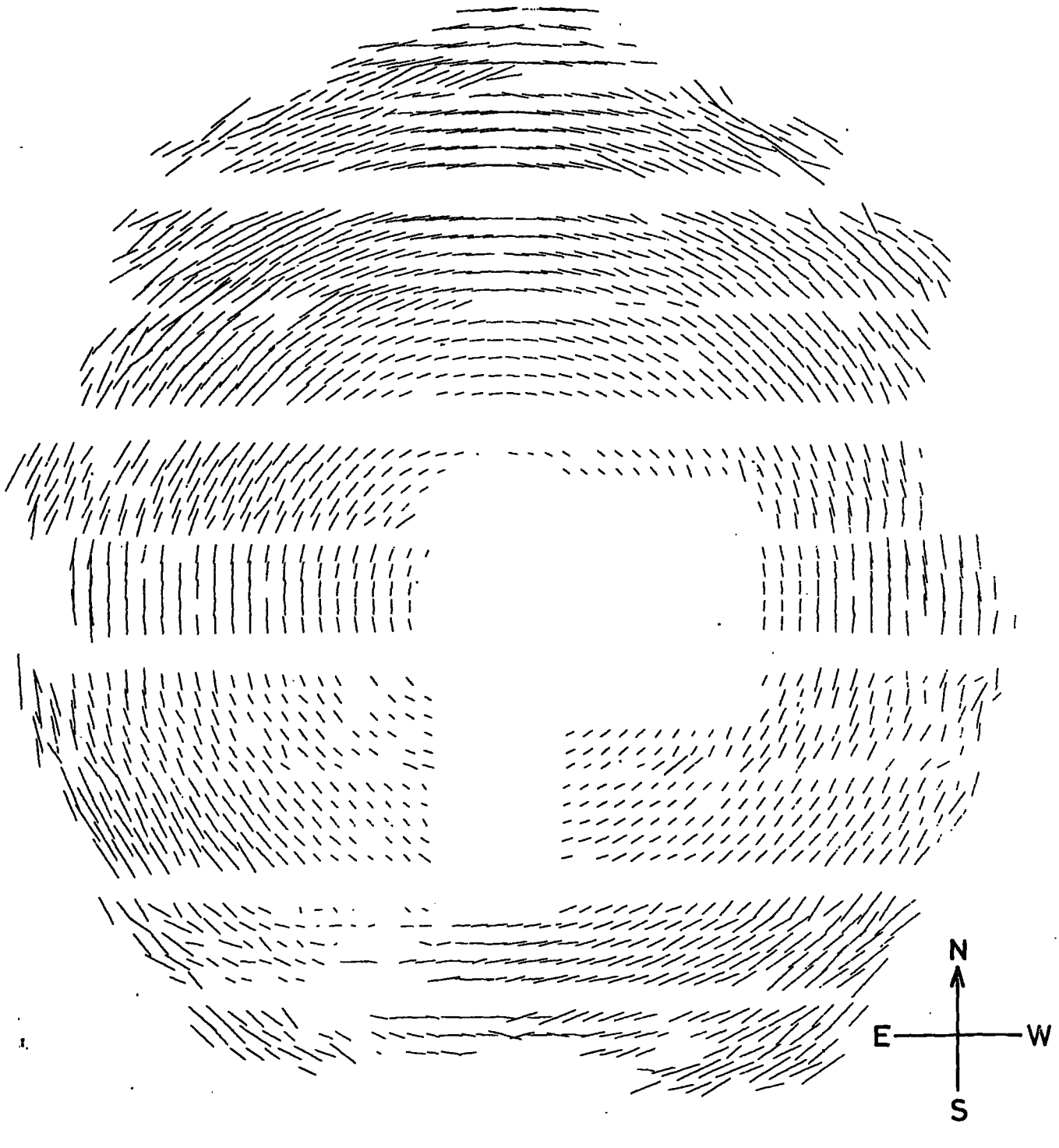


Fig.13·6 Data used when fitting the model

13.5 Results of Preliminary Modelling

13.5.1 Best-Fit Parameters

The values of the 4 model parameters which gave the best fits to the polarisation data for 5 values of the refractive index are summarised below:

Grain Type	$r_o(")$	$\gamma(^o)$	$\phi(^o)$	$a_o(\mu m)$
Biological	77.7	27.6	45.5	0.377
Ice	61.9	31.6	40.1	0.300
Silicates	69.0	46.8	32.0	0.207
Graphite	93.0	47.9	35.0	0.264
Iron	127.7	56.3	28.4	0.230

Table 7

The quality of the fits obtained can be judged by direct comparison with the data (Figures 13.7 and 13.8).

Two other grain models were also tried; a "dirty ice" model using a slightly aborbing ice-like refractive index ($m = 1.33 - 0.05i$) which gave identical results to the "clean-ice" model above and a grain mixture model containing iron, silicates and graphite proposed by Wickramasinghe and Nandy (1972) to explain the interstellar extinction law. This latter model proved to be far too polarising for any combination of the geometric parameters but, in view of its complexity, no attempt was made to adjust the grain sizes to obtain better agreement.

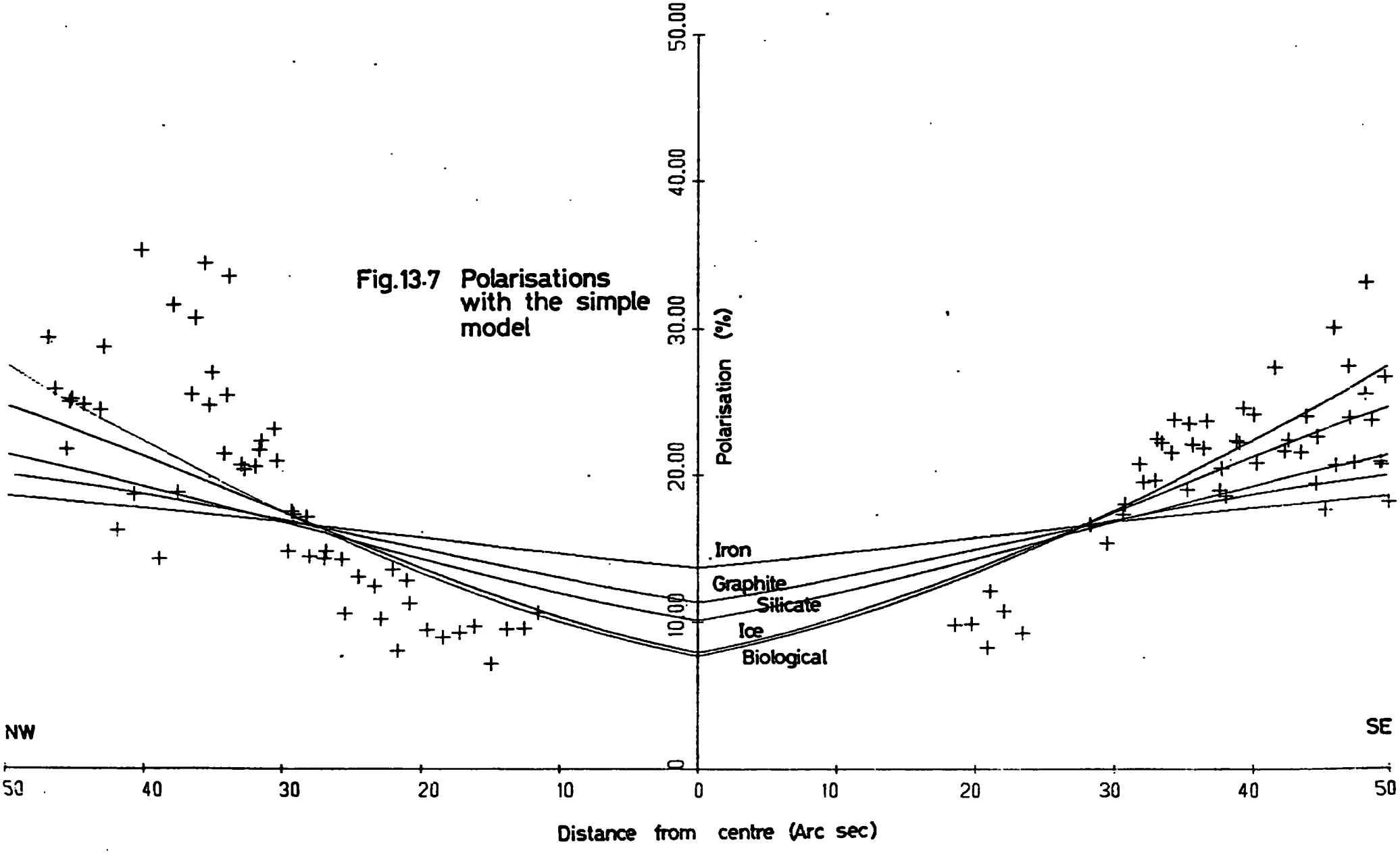


Fig.13.7 Polarisations with the simple model

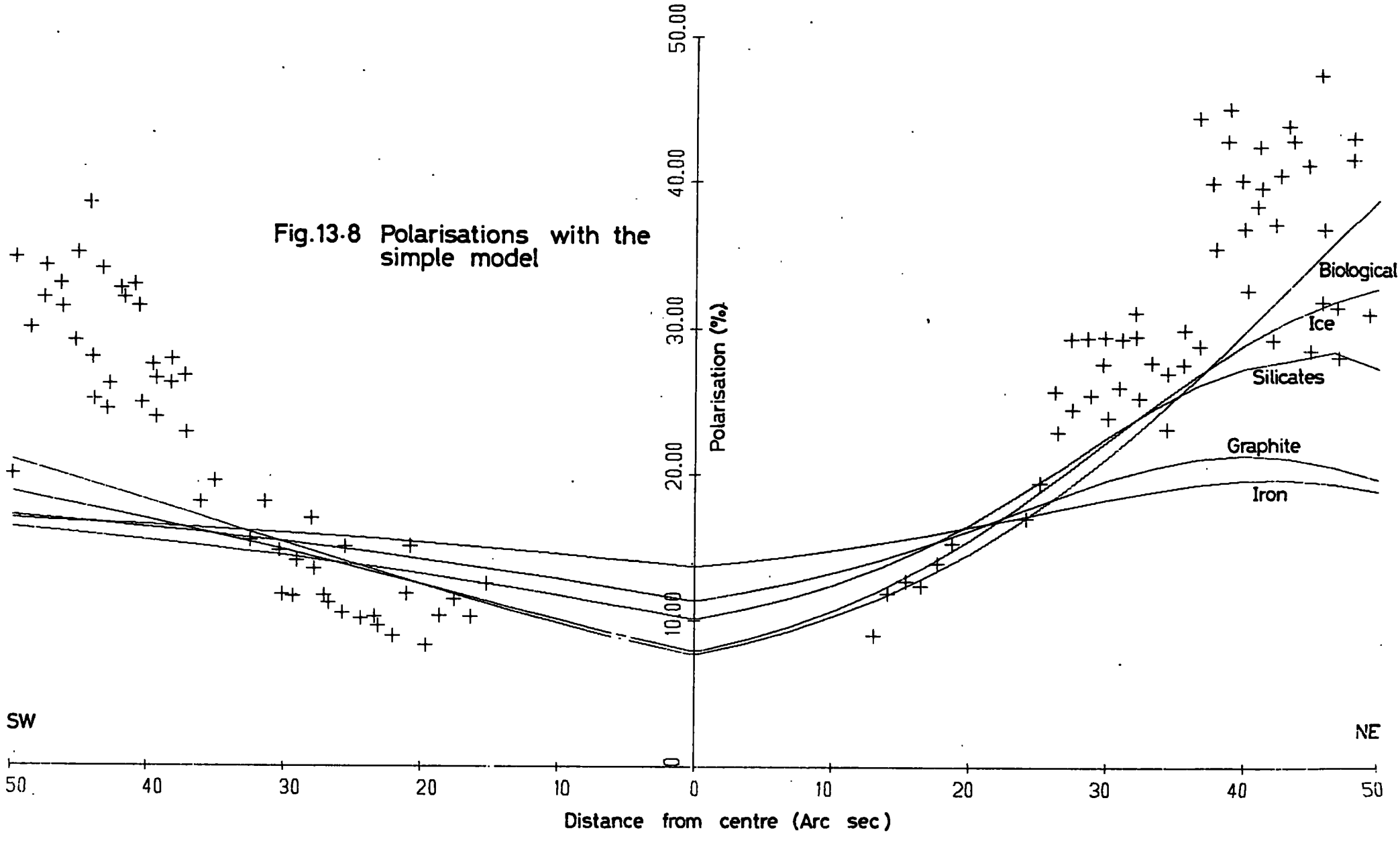


Fig.13-8 Polarisations with the simple model

13.5.2 Polarisations and Surface Brightness

The fits obtained with the parameters of table 7 are shown in figures 13.7 (NW - SE direction) and 13.8 (SW - NE direction) together with the polarisation data of figures 10.9 and 10.11 in these directions. None of the fits fully explains the data, although as expected, the smaller refractive index particles are more successful in this respect. In view of the simplifying assumptions made in this model, however, the results are sufficiently encouraging to warrant further refinement of the model.

The surface brightness resulting from this model is shown in figure 13.9 (in the SW - NE direction) assuming a grain density of 0.01 m^{-3} and a nebular distance of 470 pc. Also shown are the surface brightness data corresponding to the polarisation measurements of figure 13.8. These have been placed on an absolute scale using the B waveband measurements of Brück (1974) together with her colour measurement of the inner nebulosity (B-V = 0.1). It can be seen that the asymmetry necessary to account for the polarisation measurements does not greatly affect the surface brightness.

13.6 Internal Extinction

13.6.1 Dust Density

To explain the observed surface brightness measurements the dust density, N, must be adjusted to obtain agreement

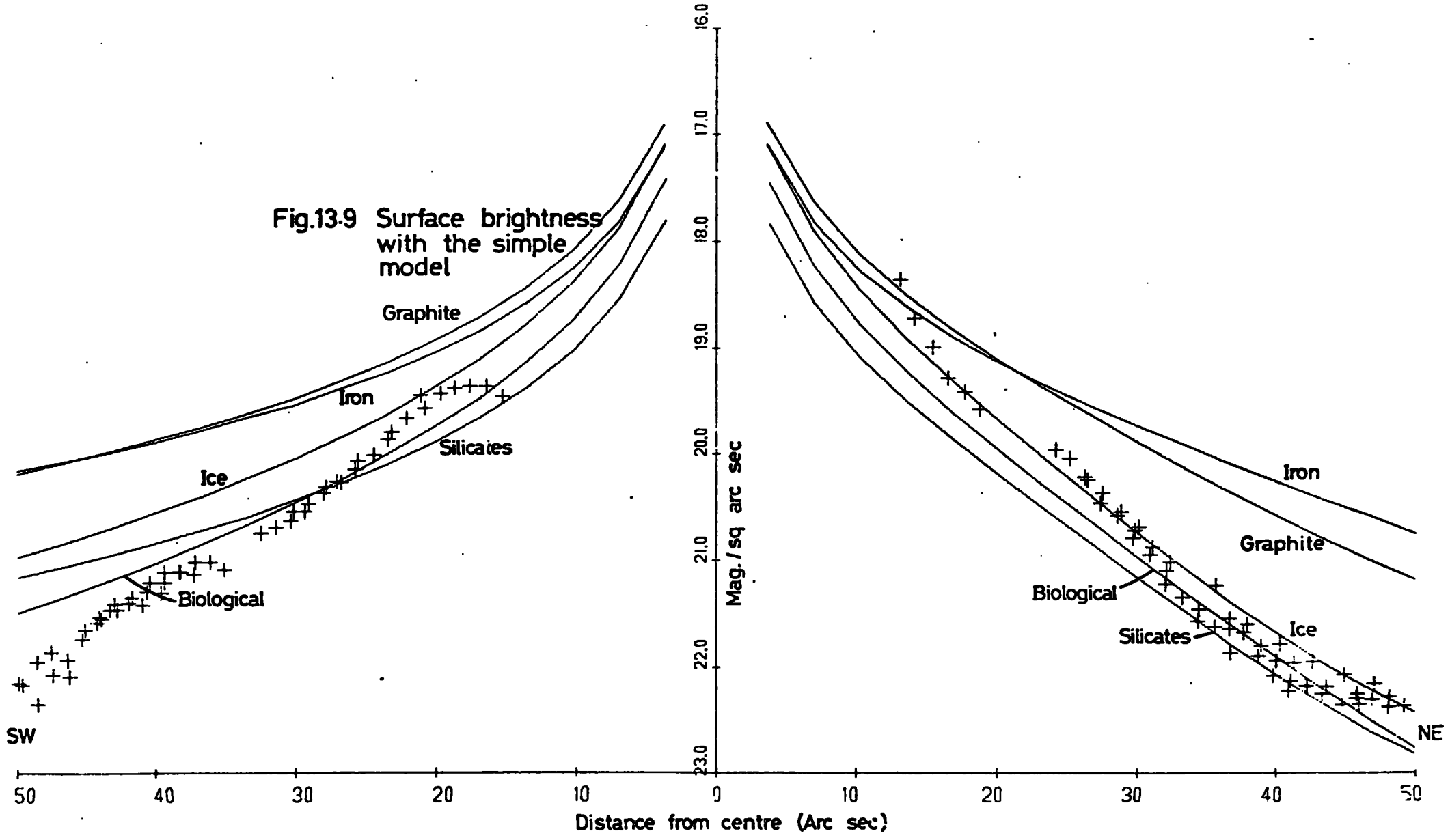


Fig.13.9 Surface brightness with the simple model

with the data of figure 13.9. The dust density, in conjunction with the optical properties of the dust necessary to explain the polarisation data, will define the extinction length, t_o, within the nebula and hence determine the surface brightness gradient. Thus, if the optical properties cannot also explain the internal extinction, no value of N will be able to explain both the surface brightness and its gradient. On this basis it may be possible to distinguish between possible dust models.

In the presence of internal extinction, equation 12.16 and the extinction factor of equation 12.18 can be used to calculate the expected polarisation. The surface brightness can be found from equation 12.8 by the inclusion of the extinction factor to give:

$$m_{NEB} = m_* - 1.086\,\frac{r_o}{t_o} - 2.5\log_{10}\left\{\frac{\lambda^2 D^2 N}{8\pi^2 r a^2}\int_{\theta_2}^{\pi}\left[F_1(\theta)+F_2(\theta)\right]\gamma(\theta)\,d\theta\right\} \qquad 13.6$$

The second term in this expression is a correction to allow for the stellar extinction due to the dust between the star and observer. The extinction length, t_o, is given by equation 12.15.

The density, N, has been adjusted in each of the best-fit cases of table 7 to give a surface brightness of 19.8 mag. per sq. arc sec. at a distance 20" from the star in a NW direction, and the resulting surface bright-

ness throughout the nebula calculated from equation 13.6.

13.6.2 Brightness with Extinction

The results of the above calculations are shown in figure 13.10 for the SW - NE direction and the required dust density and extinction factors are summarised below:

Grain Type	N (m^{-3})	t_o (")	Stellar extinction (mag.)
Biological	0.017	47.1	1.8
Ice	0.014	36.0	1.9
Silicates	0.025	25.7	2.9
Graphite	0.013	13.7	7.4
Iron	0.014	16.6	8.4

Table 8

Since the inclusion of internal extinction also affects the model polarisations this no longer represents an optimised fit to the data. However, from the tabulated stellar extinctions and the curves of figure 13.10, it is clear that graphite and iron dust are far too absorbing to account for the data. Since the effects of multiple scattering are likely to be small for these grains on account of their low albedo (0.48 for graphite and 0.61 for iron) there is no simple way of adjusting the surface brightness further to give agreement with the observations. Consequently these grain models have been rejected

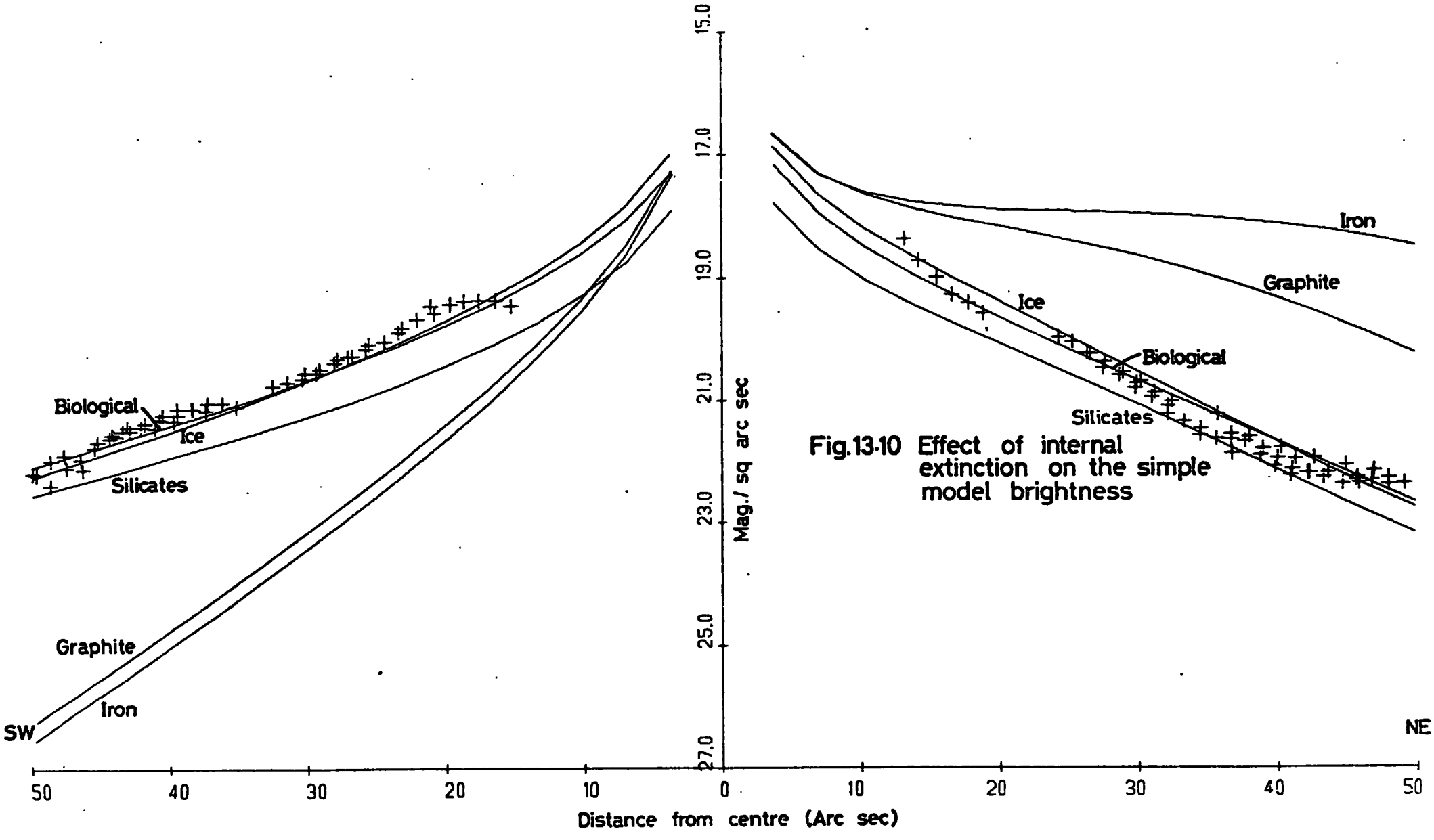


Fig.13.10 Effect of internal extinction on the simple model brightness

at this point as being too absorbing. The remaining 3 grain types give acceptable fits at this stage, however.

13.7 Finite Bandwidth Effects

13.7.1 Wavelength Integration

Until now it has been assumed that the polarisation observations were made at a single wavelength of 0.5μm. In fact, however, the optical bandpass used was rather broad; the measured filter transmission, multiplied by the spectral response of the McMullan camera, is shown in figure 13.11. Because of this broad bandwidth the effective wavelength of observation may change if the nebular colour is not constant (This is not precluded by Brück's reported constant B-V colour since her measurements have a resolution of only $\pm$ 0.25 mag. and extend to only $\sim$30" from the star) resulting in changes in polarisation equivalent to changing the mean grain size.

An approximate method of performing the integration of the model light over the passband of figure 13.11 must be used if realistic amounts of computer time are to be used. To achieve this, the passband has been represented by 2 gaussian profiles joined by a straight line section (Dotted line of figure 13.11) and the nebular light by a parabolic spectrum joining 3 points sampled within the passband. The integration can then be performed analytically and is equivalent to a weighted sum of the 3 samples. It

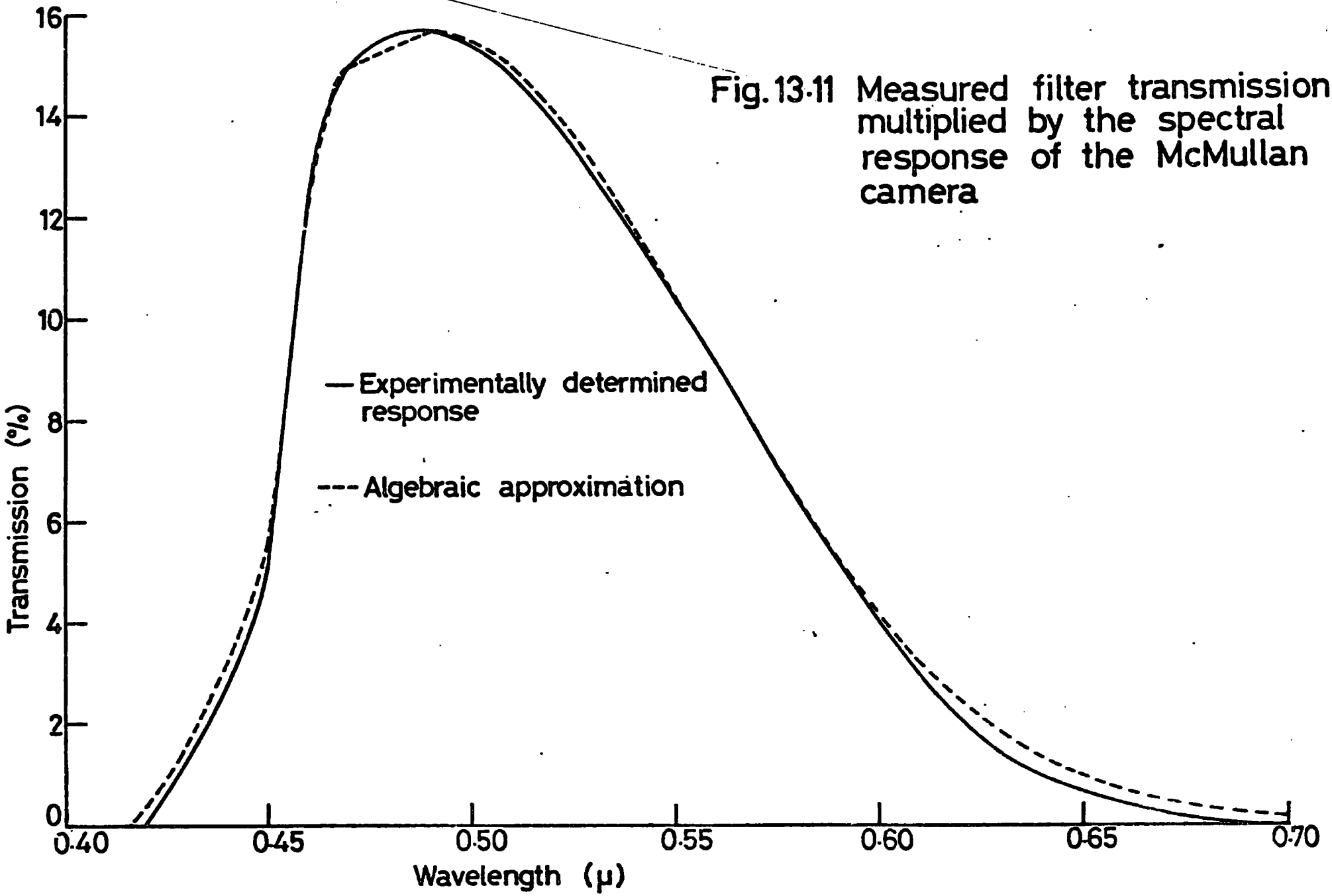


Fig. 13·11 Measured filter transmission multiplied by the spectral response of the McMullan camera

is clear, however, that many complications could be avoided in future work by the use of a narrower bandwidth so eliminating the need for this integration.

The stellar brightness at the 3 wavelengths used has been found using the spectrophotometry of Garrison and Anderson (1978b) normalised to Brück's measured V = 10.6 mag. The derived magnitudes have then been converted to a spectral energy curve and interpolated at the required wavelengths. This complete model, including both internal extinction and the wavelength integration, has been re-optimised to fit the polarisation data using the criterion of equation 13.5 with the dust density adjusted to give the observed brightness of 19.8 mag. per sq. arc sec., 20″ to the NW of V380.

13.7.2 Fits with the Complete Model

The resulting fits to the polarisation data are shown in figures 13.12 (NW - SE) and 13.13 (SW - NE) and the resulting surface brightness in the SW - NE direction is shown in figure 13.14. The model parameters giving these optimised fits are summarised below:

Grain Type	r_o(″)	γ(°)	ϕ_o(°)	a_o(μm)	N(m^{-3})
Biological	65.1	17.0	47.2	0.324	0.029
Ice	45.4	17.8	40.5	0.265	0.027
Silicates	65.4	52.2	37.6	0.174	0.043

Table 9

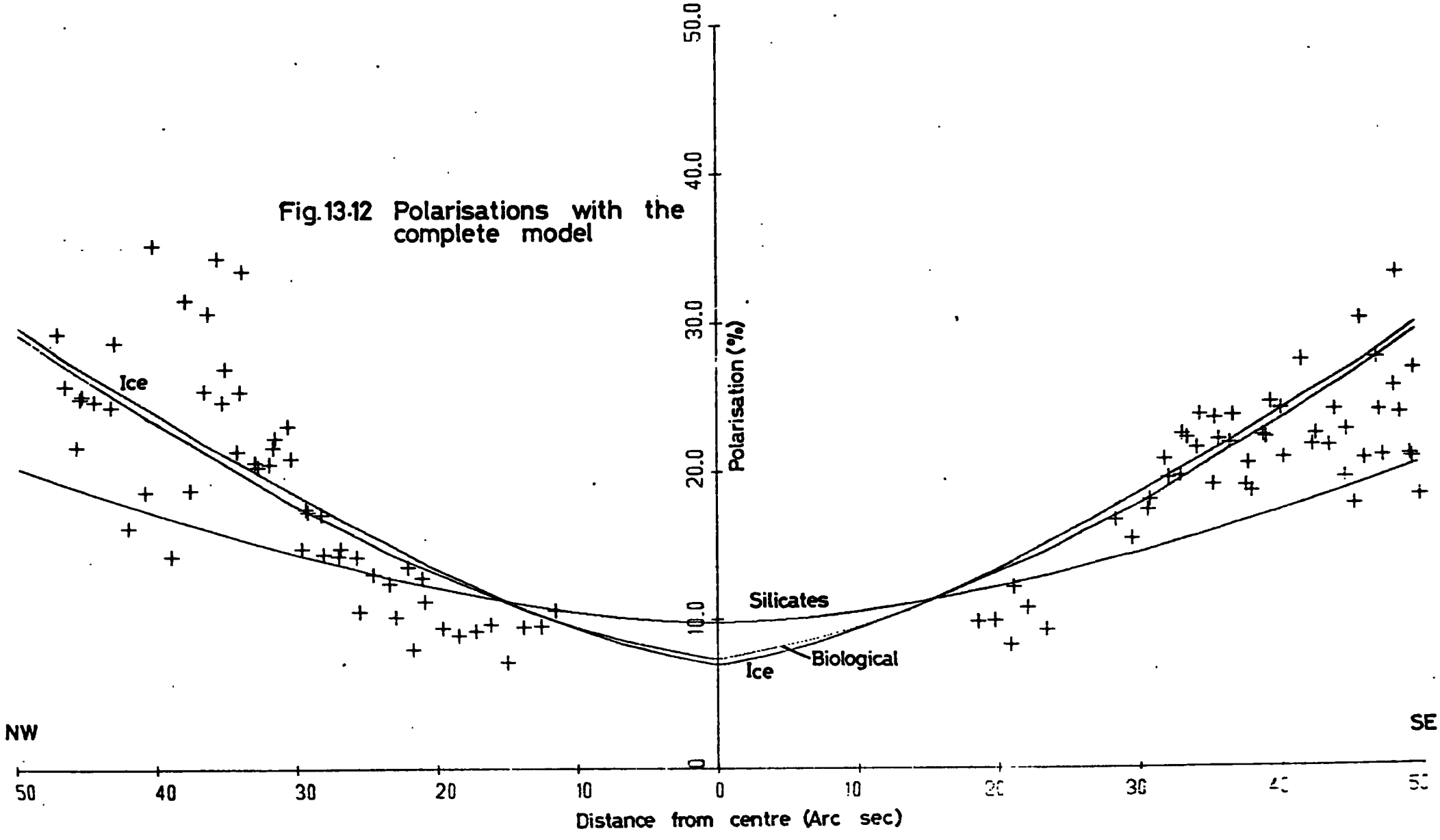


Fig. 13.12 Polarisations with the complete model

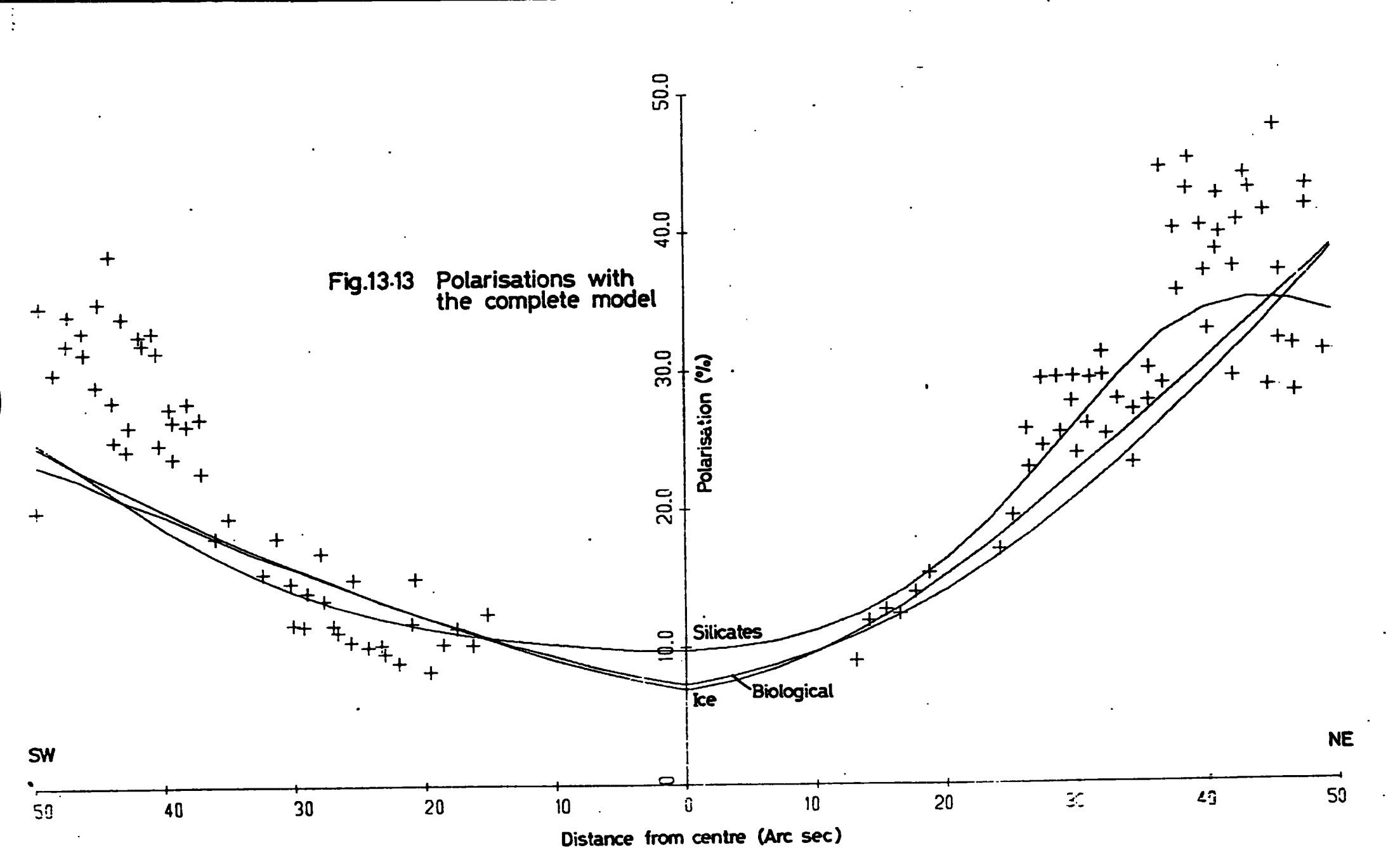


Fig.13-13 Polarisations with the complete model

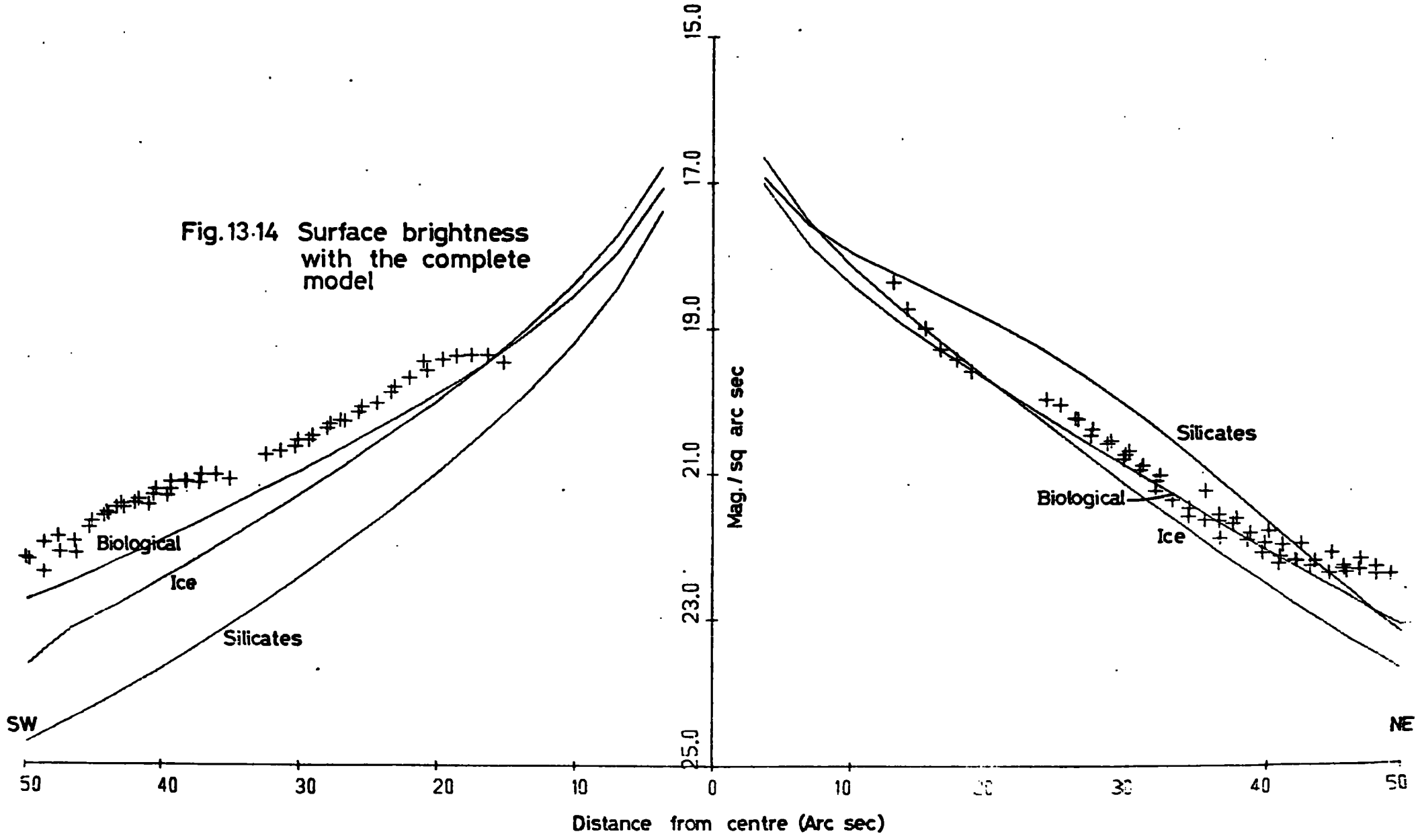


Fig. 13.14 Surface brightness with the complete model

The fits with the "biological" and ice grain models may be considered satisfactory since it is clear that small deviations from ideal geometry can account for any deficiency. For the silicates model, however, the range of polarisation observed cannot be reproduced due to the limited polarising range of the silicate grains (12.6) and it appears unlikely that any modification to the geometry would enable the highest polarisations of ~40% to be matched by this model. The surface brightness data is matched approximately by all 3 models, although the silicates model does show somewhat more asymmetry than is seen in the data. The effects of multiple scattering, however, should be taken into account before silicate grains are rejected on this basis.

Using the best-fit parameters of table 9, the following nebular quantities result:

Grain Type	t_o(")	Stellar Extinction (mag.)
Biological	53.2	1.3
Ice	31.4	1.6
Silicates	31.7	2.2

Table 10

CHAPTER 14

A MONTE-CARLO APPROACH TO MULTIPLE SCATTERING

14.1 Introduction

It is clear from the extinction lengths calculated on the basis of the best-fit models of the previous chapter (Table 10), that a significant proportion of the light propagating in the nebula must undergo at least one scattering. This, however, is not the whole story, for, except in the case of perfectly absorbing dust, the extinguished light will be at least partially re-emitted by the grains as scattered radiation. This scattering light can then be re-scattered by other grains, ad infinitum, until it finally escapes from the nebula or is attenuated to negligible levels by absorption - it is the escaped light which the observer sees.

In accounting only for the extinction within the nebula, we have assumed that all the extinguished light is lost by absorption on its first scattering. Hence, any model which takes account of multiple scattering will predict a brighter nebula than if extinction alone is taken into account. Perhaps a more important effect, however, occurs because scattered light does not form part of the radially propagating radiation field and hence does not contribute to the centro-symmetric pattern of polarisation seen in optically thin nebulae. The mis-aligned polarisation due to the multiply scattered light

results in cancellation and depolarisation when observed from outside the nebula.

14.2 Previous Work on Multiple Scattering

The problem of multiple scattering, or generally, of radiative transfer in an optically thick medium, has received theoretical treatment only in highly simplified cases. In particular, the effects of polarisation in the scattered light are usually neglected and highly idealised scattering functions, highly symmetric geometries and idealised optical depths used. Such work has been primarily applied in the theory of stellar structure and planetary atmospheres. A number of approximate techniques have also been developed in this context which allows the more useful cases involving scattering from spherical "Mie" particles to be included. A discussion of some of these techniques has been included in an article on planetary atmospheres by Coffen and Hansen (1974), who also give references to more detailed descriptions.

Recently White (1979a,b) has used one of these techniques, that of "optical depth doubling", to include the effects of multiple scattering in the case of a star illuminating a plane parallel dust slab at some distance from it. While the results are undoubtedly of great value in indicating the type of effects to be expected when multiple scattering is included, the technique is applicable

only to plane parallel incident radiation and is thus unsuited to the (presumably common) situations where the illuminating star is close to, or contained within, the nebula. Similar limitations of technique make the case of a nebula of general shape, illuminated from within, insoluble by the methods commonly used in planetary atmosphere studies.

With the lack of analytical or simple approximate techniques applicable to most reflection nebulae, a number of workers have turned to "Monte-Carlo" methods which trace the paths of imaginary "photons" through the scattering medium until they emerge from the nebula. (Such methods have also found application in planetary atmosphere studies.)

Monte-Carlo methods have achieved notoriety for requiring excessive amounts of computation time, since the results obtained have a statistical error which varies like $[\text{computer time}]^{-\frac{1}{2}}$, so that long computer run times are required if high accuracy is needed. In reflection nebulae, a further measure of inefficiency is introduced since, because of the random nature of the light scattering process, there is no way to know in advance where a given trial "photon" will emerge from the nebula. If it does not emerge in the direction of the observer, it, together with the computation time involved, is lost. Since the observer subtends only a small solid angle at the nebula, most photons used are lost in this manner, resulting in very low efficiency. To overcome this problem

cases of high symmetry, where all emerging directions are equivalent, have tended to dominate investigations using Monte-Carlo techniques. Despite the problems involved, however, the "last resort" Monte-Carlo method has been applied with some success in a number of cases involving light scattering in nebulae. In nearly all of these the effects of polarisation have been neglected (e.g. Mattila, 1970; Roark, Roark and Collins, 1974; Witt and Stephens, 1974; Fitzgerald et Al., 1976; Andriesse et Al., 1977)

Witt (1977a,b,c,d), in a series of papers, has treated the case of plane parallel reflection nebulae with finite optical depth. This work, which neglects polarisation effects and uses an idealised analytic scattering function, overcomes the problem of low symmetry by integrating the results over large areas of the nebular surface and by using large acceptance angles for the emergent photons. (i.e. The observer subtends a large solid angle at the nebula.) Nontheless, computational efficiency is still low and the results are accurate enough only for the total light intensity, not the polarisation.

Very little work has been done using Monte-Carlo techniques to determine polarisations in reflection nebulae. The computation time is likely to be very long in this context, since higher relative accuracies are required than for calculations involving only intensity - typically a 10 fold increase in accuracy, or a 100 fold increase in

computer time is required. Roark, Roark and Collins (1974) in a Monte-Carlo study of plane-parallel reflection nebulae used a number of photons, n, in the range 6,000,000 to 30,000,000, achieving average accuracies of $\pm$ 7% in intensity. The necessary increase in computation time to enable polarisation effects to be included in their model is clearly excessive.

The most relevant work to the present study was undertaken by Vanysek and Solc (1973). They considered the case of a sphere containing a constant density of dust, with a star at its centre, and considered the polarisation as a function of offset distance for a variety of optical depths. Their scattering functions were based on a size distribution of silicate spheres as regards total scattered intensity, but on a single sized sphere as regards polarising properties - a choice which makes comparison with other results virtually impossible. The depolarising effect of multiple scattering was, however, conclusively demonstrated. While no errors are given for the results presented, a graphical display shows considerable dispersion although the full advantage of spherical symmetry was used.

14.3 The New Approach

It is clear from the above that the untreated case of polarisation in reflection nebulae without high symmetry cannot be approached simply by using more computer time and relying on the $1/\sqrt{n}$ convergence of Monte-Carlo results.

Any increase in accuracy must rely on improvements to the inherent computational efficiency of the Monte-Carlo technique.

The most important contribution to inefficiency is the loss of all photons which do not emerge from the nebula in the direction of the observer. The new approach to be described ensures that nearly all the photons considered contribute to the result (several times in fact), giving an immediate increase in efficiency by a very large factor. In addition, a number of variance reducing techniques, standard in Monte-Carlo work, are employed to further increase efficiency. Using this method, accuracies of $\sim^{+}_{-}$ 0.5% can usually be achieved with n = 1000.

The scattering functions used are based on a size distribution of "Mie" particles, the distribution function and refractive index (including an absorptive part if necessary) being quite general, and no simplifying approximations are necessary. The range of nebular geometries which can be considered is wide, the only restrictions being that the star is contained within the nebula and that the nebula boundary is nowhere concave. While the approach adopted here applies to constant dust density, it can, in principle, be extended to include spherically symmetric variations in density by the use of spherical shells of differing dust density as have been used by other workers (e.g. Witt and Stephens, 1974).

14.4 Scattering of Polarised Light

Before proceeding to describe the details of the Monte-Carlo technique, we must consider the scattering, by Mie particles, of light which is already polarised. This does not arise in the case of single scattering because initially unpolarised light can only acquire linear polarisation in a single scattering from spherical particles. However, subsequent scatterings of linearly polarised light give rise to circular and elliptical polarisations, so in general we need to consider the scattering of light with a general state of polarisation.

With reference to section 11.2 and figure 11.1, the emergent light, after scattering from a single sphere, has complex amplitudes B_1 and B_2 given by:

$$(B_1, B_2) = \begin{bmatrix} S_1(\theta) & 0 \\ 0 & S_2(\theta) \end{bmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \qquad 14.1$$

where A_1, A_2 refer to the incident light.

As discussed in section 1.4, from the principle of optical equivalence, we need consider only the four Stokes parameters of the light to fully define its state of polarisation. Using the definitions of the Stokes parameters, and equation 14.1, it can be shown (e.g. van de Hulst, 1957) that the Stokes parameters of the scattered light (I_s, Q_s, U_s, V_s) are related to those of the incident light (I_i, Q_i, U_i, V_i) by the matrix equation:

$$(I_s, Q_s, U_s, V_s) = \begin{bmatrix} C & D & 0 & 0 \\ D & C & 0 & 0 \\ 0 & 0 & E & -F \\ 0 & 0 & F & E \end{bmatrix} \cdot \begin{bmatrix} I_i \\ Q_i \\ U_i \\ V_i \end{bmatrix}$$

14.2

where the matrix elements are real quantities given by:

$$C = \frac{1}{2}\left(\frac{\lambda}{2\pi}\right)^2 \left[|s_1(\theta)|^2 + |s_2(\theta)|^2\right] \qquad 14.3$$

$$D = \frac{1}{2}\left(\frac{\lambda}{2\pi}\right)^2 \left[|s_1(\theta)|^2 - |s_2(\theta)|^2\right] \qquad 14.4$$

$$E = \frac{1}{2}\left(\frac{\lambda}{2\pi}\right)^2 \left[s_1(\theta)\, s_2^*(\theta) + s_2(\theta)\, s_1^*(\theta)\right] \qquad 14.5$$

$$F = \frac{i}{2}\left(\frac{\lambda}{2\pi}\right)^2 \left[s_1(\theta)\, s_2^*(\theta) - s_2(\theta)\, s_1^*(\theta)\right] \qquad 14.6$$

and $s_1(\theta)$, $s_2(\theta)$ are the complex Mie scattering amplitudes of equations 11.10 and 11.11.

When a size distribution of Mie particles is present, each size will scatter independently according to equation 14.2, and since the Stokes parameters of independent light beams are additive, it follows that the scattering matrix appropriate to a size distribution is found by summing the matrices appropriate to each particle size, weighted with the number of such particles, i.e. by setting:

$$C = \frac{1}{2}\left(\frac{\lambda}{2\pi}\right)^2 \int_0^\infty n(a) \left[|s_1(\theta)|^2 + |s_2(\theta)|^2\right] da \qquad 14.7$$

and similarly for D,E,F.

Hence, if we know the Stokes parameters for a beam of light, referred to axes defined by the scattering plane, we can find from Mie's theory the parameters of the light after scattering through any angle by a size distribution of spherical particles.

14.5 The Monte-Carlo Technique

14.5.1 Basic Principles

The techniques of Monte-Carlo simulation reduce ultimately to the problem of integration of a function over some range by using random samples, i.e. of finding:

$$R = \int_a^b f(x)\, dx \qquad 14.8$$

where f(x) may be only indirectly known or not directly integrable. Usually numerical quadrature methods are more appropriate, but this is not necessarily so if the integral is multi-dimensional, as is the case with multiple scattering.

The case of a uni-dimensional integral will be taken to illustrate 2 principles to be used in the Monte-Carlo technique to be described:

1) If r_i are a set of samples of a random variable r, with a probability distribution function (PDF) which is uniform in the interval $[a,b]$, i.e.

$$\mathbb{P}(r)\,dr = \frac{1}{|a-b|}\,I_{[a,b]}\,dr \qquad 14.9$$

then:

$$\hat{R}_1 = \frac{1}{n}\sum_{i=1}^{n} f(r_i) \qquad 14.10$$

is an unbiassed estimator of R (i.e. the expectation value of $\hat{R}_1$ is equal to R), and the statistical error on the estimate decreases like $1/\sqrt{n}$ for large values of n.

ii) If r_i are a set of samples of a random variable r, with a PDF G(r), proportional to the positive function g(r) in the interval $[a,b]$, i.e.

$$\mathbb{P}(r)\,dr = G(r)\,dr = \frac{g(r)}{\left[\int_a^b g(y)\,dy\right]}\,I_{[a,b]}\,dr \qquad 14.11$$

then:

$$\hat{R}_2 = \frac{1}{n}\sum_{i=1}^{n} \frac{f(r_i)}{G(r_i)} \qquad 14.12$$

is an unbiassed estimator of R. Furthermore, if G(r) is such that the ratio f(r) / G(r) is nearly constant over the range of integration, all the elements in the summation will have nearly the same value and the statistical error is much reduced. (Hammersley and Handscomb, 1975 -

importance sampling). Hence, by taking samples from a deliberately biassed distribution, but applying the weighting function $1/G(r)$ to correct for this, an overall increase in computational efficiency results.

14.5.2 The Line of Sight Integration

Consider the nebula of figure 14.1. At any point P, on the line of sight shown, the light will be propagating in a variety of directions with a variety of states of polarisation, dependent on the path and number of scatterings it has undergone in reaching P.

Representing this light flux as a function $F(\underline{K},\underline{S},z)$ of the propagation direction $\underline{K}$, Stokes vector $\underline{S}$ and position on the line of sight z, the light power scattered into a unit solid angle in the direction of the observer by a line of sight element dz is:

$$d\underline{R} = NA\,dz \iint_{\text{all } \underline{K},\underline{S}} F(\underline{K},\underline{S},z)\ M(\underline{K})\ \hat{\underline{S}}\ d^3K\ d^4S \qquad 14.13$$

where M is the scattering matrix of equation 14.2 referred to appropriate axes depending on the propagation direction $\underline{K}$, and $\underline{S}$ is the normalised Stokes vector ($\hat{S}_1 = 1$). The scattered power $d\underline{R}$ has the 4 components of a Stokes vector.

The scattered light is attenuated by extinction before reaching the observer, so the total light power from the entire line of sight is:

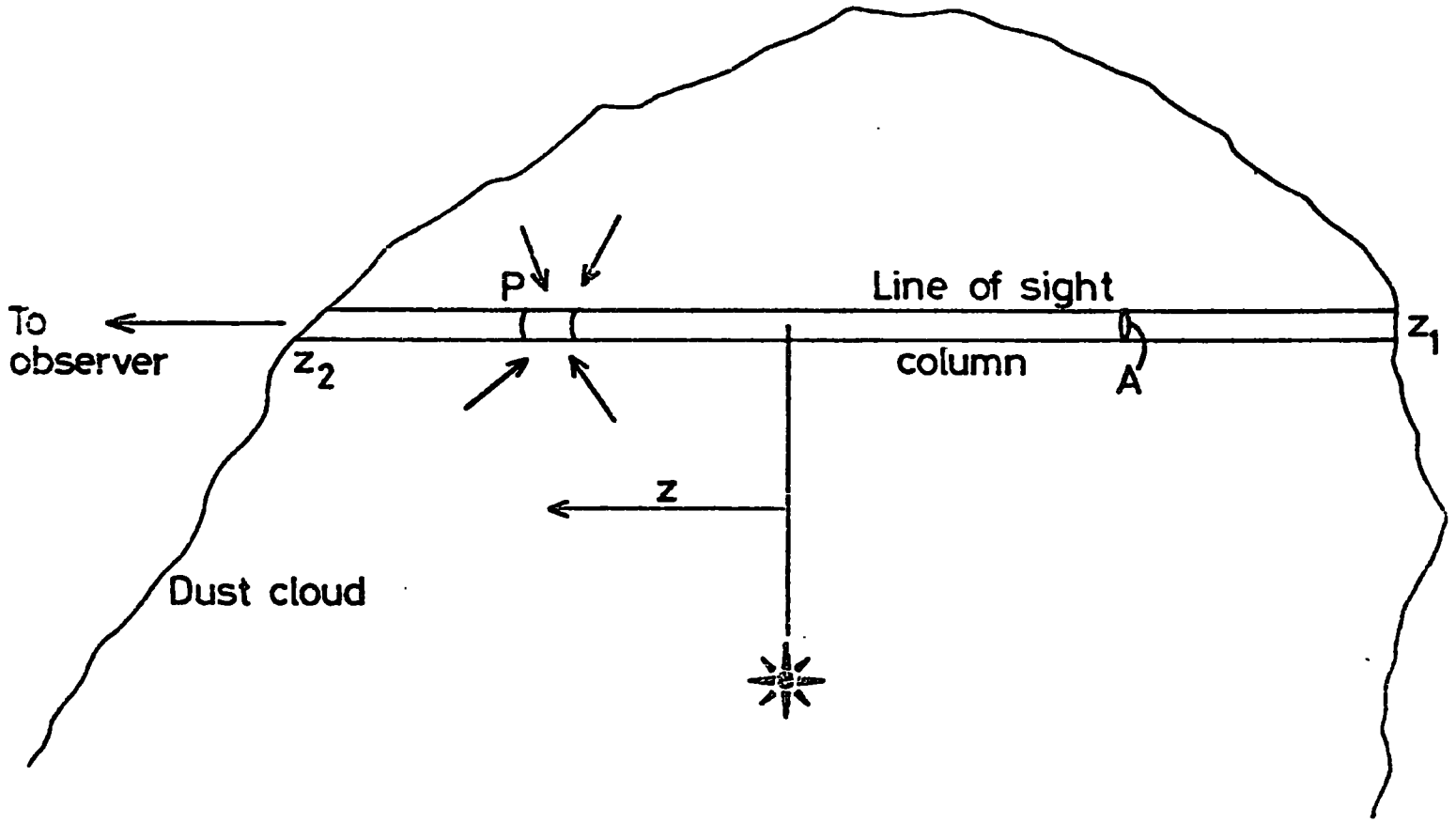


Fig.14-1 Scattering in an optically thick nebula

$$\underline{R} = NA \int_{z_1}^{z_2} \iint_{\underline{K},\underline{S}} F(\underline{K},\underline{S},z)\, M(\underline{K})\, \hat{\underline{S}}\, d^3K\, d^4S \exp - \left(\frac{z_2 - z}{t_o} \right) dz \qquad 14.14$$

Using the results of section 12.3, the nebular brightness in flux units at the observer per square arc second of nebula is:

$$\underline{F}_o = \frac{\underline{R}}{A\, \alpha^2} \qquad 14.15$$

We can estimate the integral of equation 14.14 using the Monte-Carlo technique if we can obtain random samples of $\underline{K},\underline{S},z$ with a PDF proportional to $F(\underline{K},\underline{S},z)$. This can be done by following photon paths through the nebula as will be described.

14.5.3 Sampling the Flux Distribution

For a beam of light propagating through a uniform scattering medium, the number of photons scattered per unit volume is proportional to the light flux in the beam. Hence, if we follow the paths of random imaginary photons through the nebula, the density of n th order scatterings in a particular region provides a measure of the light flux due to (n - 1)th order scatterings.

In more precise terms, if we take as our energy unit one photon and consider it propagating out from the central star, the probability of it having its n th

scattering within a volume element dV at some position $\underline{r}$ with Stokes vector in the range $\underline{S}$ to $\underline{S} + d\underline{S}$ and propagation direction in the range $\underline{K}$ to $\underline{K} + d\underline{K}$ can be written:

$$P_n\ (\underline{K},\underline{S},\underline{r})\ \ dV\ \ d^3K\ \ d^4S \qquad 14.16$$

If the energy flux at this point is $F_{n-1}\ (\underline{K},\underline{S},\underline{r})$, due to (n-1) th order scatterings, for a unit of energy emitted by the star, then the probability of a scattering occurring within a volume element dV is given by:

$$F_{n-1}\ (\underline{K},\underline{S},\underline{r})\ \ d^3K\ \ d^4S\ \ \frac{dV}{t_o} \qquad 14.17$$

and equating these two quantities gives

$$F_{n-1}\ (\underline{K},\underline{S},\underline{r}) = t_o\ P_n\ (\underline{K},\underline{S},\underline{r}) \qquad 14.18$$

Hence the total radiation flux due to all orders of scattering is given by:

$$F(\underline{K},\underline{S},\underline{r}) = t_o \sum_{n=1}^{\infty} P_n(\underline{K},\underline{S},\underline{r}) \qquad 14.19$$

Thus the required random samples from F can be obtained by following random photons and taking the values of $\underline{K},\underline{S},\underline{r}$ for the n th scattering as a sample from P_n.

14.5.4 Photon Paths in the Nebula

In following the paths of random photons through the nebula, we ascribe to them polarisation states so that the state after n scatterings can be used in calculating the results. Because the polarisation acquired at one scattering will influence the probable direction of the next scattering, generating photon paths which all have equal probability requires forming samples from PDF's composed of Mie scattering functions and the current state of photon polarisation. Such an approach is extremely inefficient computationally.

To avoid this problem, method (ii) (14.4.1) is used to sample the photon paths from an easily generated distribution and to apply weighting factors to remove the bias introduced by doing so. The weighting factors then depend on the polarisation state at each scattering, but are easily calculated.

The variable to which this is applied is the scattering angle, θ, at each scattering (the distance between scatterings is exponentially distributed and causes no problems). Some workers, in adopting this method, have chosen the scattering angles to correspond to isotropic scattering ($\cos\theta$ uniformly distributed between ± 1), and the weighting factor is then given by the scattering matrix of equation 14.2. This approach, however, introduces a large measure of inefficiency, since, because the Mie functions are so forward throwing, the photon paths generated in fact have a low real probability and thus

receive low weight. A few photons, however, which undergo forward scatterings only, receive large weights so that the statistical error on the result is dominated by these few photons and thus increased. To improve efficiency we need to ensure that nearly equal weight is given to all photons - i.e. we need to sample the scattering angles from a forward throwing distribution which approximates the Mie scattering functions.

The appearance of the log. plots of figure 11.5 suggests that a distribution of the form:

$$P(\cos\theta)\ d(\cos\theta) = \frac{k}{e^{k}-e^{-k}}\ \exp(k\cos\theta)\ d(\cos\theta) \qquad 14.20$$

might be appropriate. Samples from this distribution are easily generated from samples of γ, uniformly distributed between 0 and 1 by taking:

$$\cos\theta = \frac{1}{k}\ln\left\{\gamma(e^{k} - e^{-k}) + e^{-k}\right\} \qquad 14.21$$

If k is chosen so that:

$$k = \frac{1}{2}\ln\left\{\frac{F_1(180^{o}) + F_2(180^{o})}{F_1(0) + F_2(0)}\right\} \qquad 14.22$$

then the forward: backward scattering ratio will precisely match that of the Mie functions.

14.5.5 Generating the Photon Paths

We will now consider how to generate photon paths in a nebula which extends infinitely in all directions from the origin of (x,y,z) coordinates. (In what follows, $\mathcal{R}_j$ is a sample from a PDF uniform between $\pm$ 1 and τ_j is a sample from a PDF uniform between 0 and 1)

i) The initial direction of photon propagation is random, and its propagation vector $\underline{v}_1$ is found from:

$$v_x = \mathcal{R}_j \qquad 14.23$$

$$v_y = \sqrt{1 - v_x^2} \sin (\pi . \mathcal{R}_{j+1}) \qquad 14.24$$

$$v_z = \sqrt{1 - v_x^2} \cos (\pi . \mathcal{R}_{j+1}) \qquad 14.25$$

This procedure generates 3 components of a randomly orientated unit vector from 2 random numbers $\mathcal{R}_j, \mathcal{R}_{j+1}$.

ii) The distance to the first scattering, d, is exponentially distributed, with a mean path length t_o, and is found from:

$$d_1 = - t_o \ln (\tau_j) \qquad 14.26$$

The position of the first scattering is thus:

$$\underline{r}_1 = d_1 . \underline{v}_1 \qquad 14.27$$

iii) We now need to generate a new propagation direction $\underline{v}_2$, with $\cos\theta$ distributed as in equation 14.20,

but with the azimuthal scattering angle uniformly distributed (Figure 14.2).

Choosing cos θ according to equation 14.21, we have to find 2 other components of $\underline{v}_2$ normal to $\underline{v}_1$. Generating 2 orthogonal unit vectors from:

$$\underline{n}_1 = \frac{\underline{v}_1 \times \underline{p}}{|\underline{v}_1 \times \underline{p}|} \qquad 14.28$$

where $\underline{p}$ is any vector not parallel to $\underline{v}_1$, and:

$$\underline{n}_2 = \underline{v}_1 \times \underline{n}_1 \qquad 14.29$$

The required new direction $\underline{v}_2$ can be written:

$$\underline{v}_2 = \underline{v}_1 \cos\theta + \sin\theta \left[\underline{n}_1 \cos(\pi.\mathcal{R}_j) + \underline{n}_2 \sin(\pi.\mathcal{R}_j)\right] \qquad 14.30$$

which is clearly normalised to $|\underline{v}_2| = 1$

iv) The position of the 2nd scattering is found by selecting another distance d from equation 14.26 and finding:

$$\underline{r}_2 = \underline{r}_1 + d_2 \underline{v}_2 \qquad 14.31$$

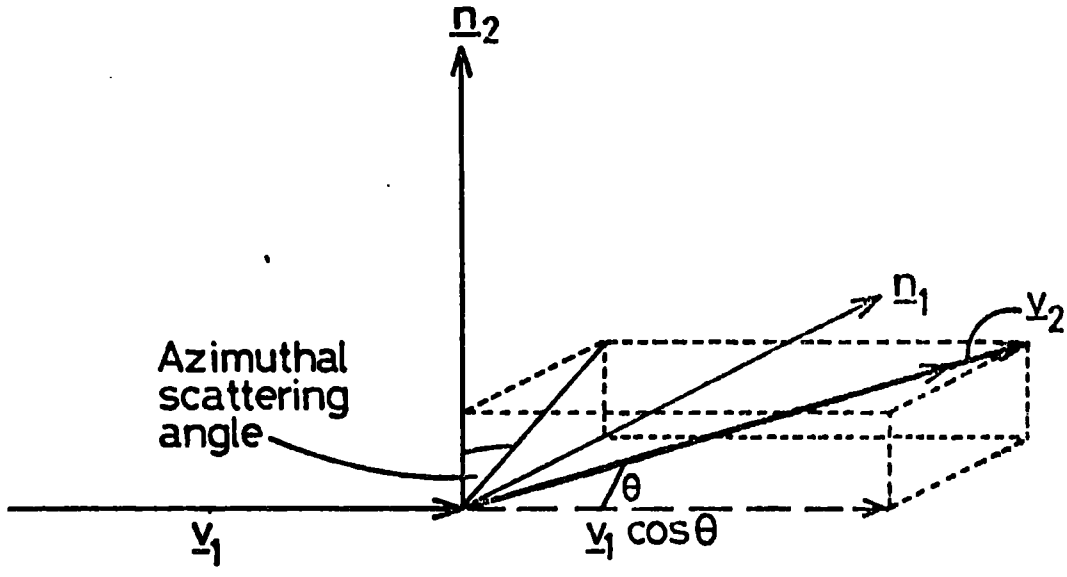


Fig.14.2 Generating a new photon direction

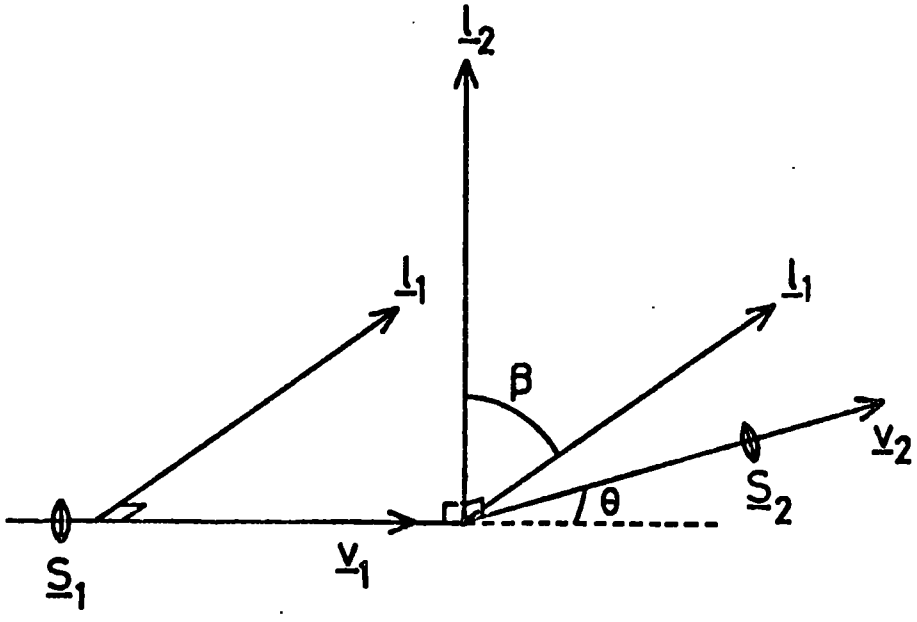


Fig 14.3 The important directions during scattering

The process is then repeated to generate a path with as many scatterings as required.

14.5.6 Polarisation along the Photon Path

We now need to determine the state of polarisation resulting if an initially unpolarised photon follows one of the above random paths. The state of polarisation between any two scatterings is represented by a Stokes vector with 4 components, referred to a reference direction $\underline{\ell}$, perpendicular to the propagation direction. Initially, the Stokes vector is (1,0,0,0) and $\underline{\ell}$ may be in any direction. After subsequent scatterings, however, $\underline{\ell}$ needs to be know to define the polarisation state.

Consider the scattering of figure 14.3. We first generate a new reference vector $\underline{\ell}_2$, perpendicular to the scattering plane, from:

$$\underline{\ell}_2 = \frac{\underline{v}_1 \times \underline{v}_2}{|\underline{v}_1 \times \underline{v}_2|} \qquad 14.32$$

The Stokes vector must be referred to this direction to compute the result of the scattering, hence we determine β, the angle between the old and new reference vectors from:

$$\cos\beta = \underline{\ell}_1 \cdot \underline{\ell}_2 \qquad 14.33$$

and

$$\sin\beta = (\underline{\ell}_1 \times \underline{\ell}_2).\underline{v}_1 \qquad 14.34$$

The rotated Stokes vector is then found from the transformation:

$$I' = I \qquad 14.35$$

$$Q' = Q \cos 2\beta + U \sin 2\beta \qquad 14.36$$

$$U' = U \cos 2\beta - Q \sin 2\beta \qquad 14.37$$

$$V' = V \qquad 14.38$$

together with:

$$\cos 2\beta = 1 - 2 \sin^2 \beta \qquad 14.39$$

$$\sin 2\beta = 2 \sin \beta \cos \beta \qquad 14.40$$

The Stokes vector after scattering can now be found using the matrix of equation 14.2 as:

$$(I'', Q'', U'', V'') = w(\theta)\, b_o \begin{bmatrix} C & D & 0 & 0 \\ D & C & 0 & 0 \\ 0 & 0 & E & -F \\ 0 & 0 & F & E \end{bmatrix} \begin{pmatrix} I' \\ Q' \\ U' \\ V' \end{pmatrix} \qquad 14.41$$

where the weight function $w(\theta)$ is derived from the PDF of $\cos \theta$ used, divided by the PDF corresponding to isotopic scattering, i.e:

$$w(\theta) = \left[\left(\frac{2k}{e^{k} - e^{-k}} \right) \exp (k \cos \theta) \right]^{-1} \qquad 14.42$$

The constant, b_o, represents a normalisation of the scattering matrix such that, with no absorption, the mean scattered intensity equals the mean incident intensity (no photons are lost in the scattering process). This requires:

$$\int_0^{\pi} b_o \, C(\theta) \sin\theta \, d\theta = \int_0^{\pi} \sin\theta \, d\theta \qquad 14.43$$

which is achieved by setting:

$$b_o = \frac{4\pi}{G_{EXT}} \qquad 14.44$$

In the case of absorbing grains, this normalisation is still valid, although the scattered photons then have a lower mean intensity as a result of the absorption loss.

Having thus determined the Stokes vector of the scattered light together with a new reference direction $\underline{\ell}_2$, we can proceed to the next scattering point and repeat the process and thus determine the polarisation state at any point on the photon's path.

14.5.7 Sampling the Flux on the Line of Sight

Considering, still, the case of an infinite nebula: having determined a given photon path and the corresponding polarisation state after the n th scattering, this (together with the weight function which is included in the Stokes

vector) represents a sample from the PDF $P_n(\underline{K},\underline{S},\underline{r})$ of equation 14.19 and we can use a set of these samples to estimate the multiple scattering integral.

The position of the n th scattering is, however, random and will not generally lie on the required line of sight. Because of the spherical symmetry, however, if the entire photon path is rotated until the n th scattering <u>does</u> lie on the line of sight, it is still a valid photon path. If the probability of this scattering lying in an element $d\gamma$ of the line of sight column (Figure 14.4) is:

$$P'_n \ (\underline{K},\underline{S},\gamma) \ d\gamma \ d^3K \ d^4S \qquad 14.45$$

then this equals the probability of it having occurred within a shell of thickness dr, which is:

$$P_n \ (\underline{K},\underline{S},\underline{r}) \ 4\pi r^2 \ dr \ d^3K \ d^4S \qquad 14.46$$

Hence:

$$P'_n \ (\underline{K},\underline{S},\gamma) = P_n \ (\underline{K},\underline{S},\underline{r}) \ 4\pi r^2 \left|\frac{d\gamma}{dr}\right| \qquad 14.47$$

Thus by rotating the photon path, we produce samples from P'_n, related to P_n by equation 14.47. To compensate for this, we simply introduce the geometrical weighting factor:

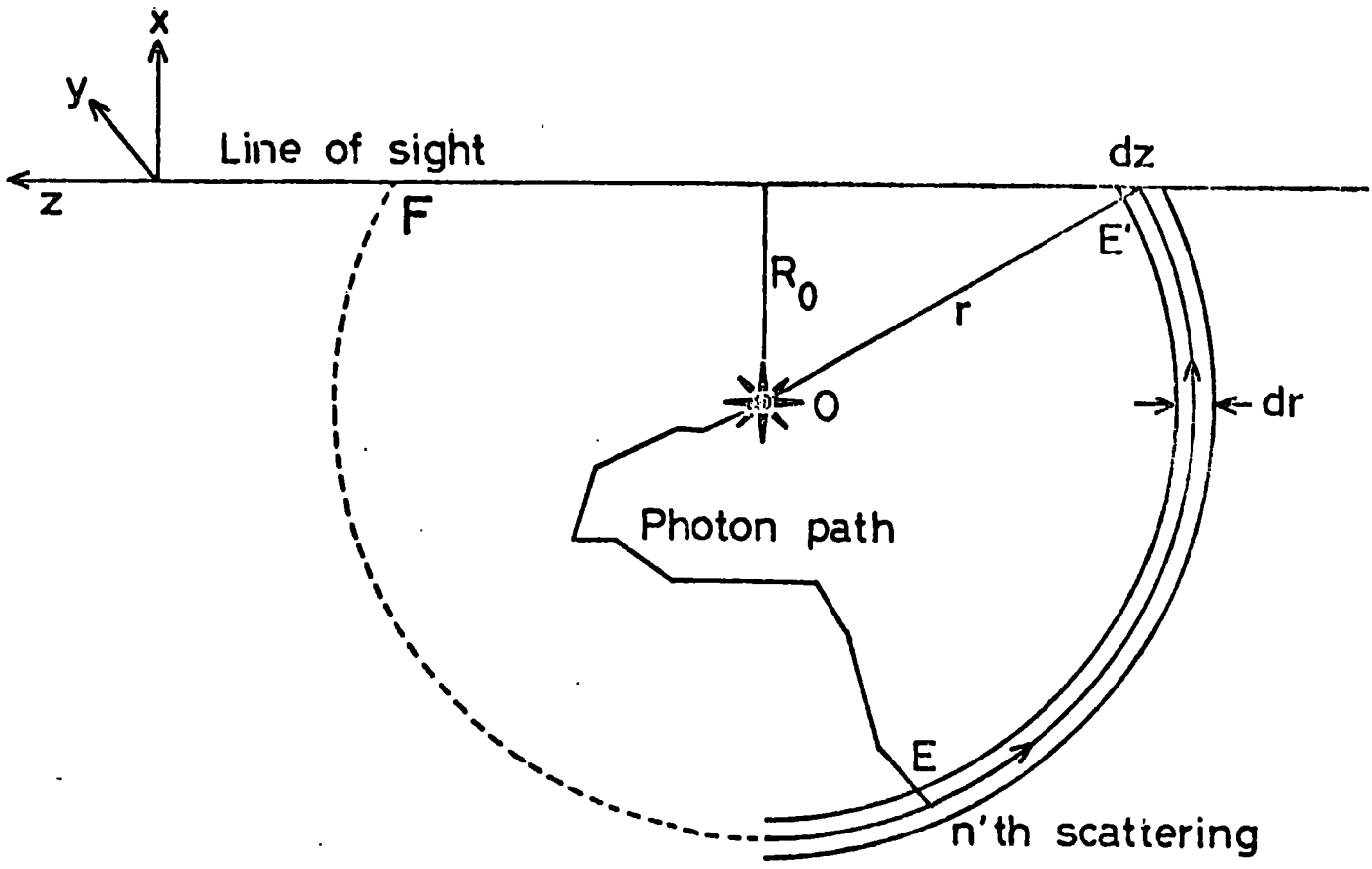


Fig.14.4 Rotating the photon path on to the line of sight

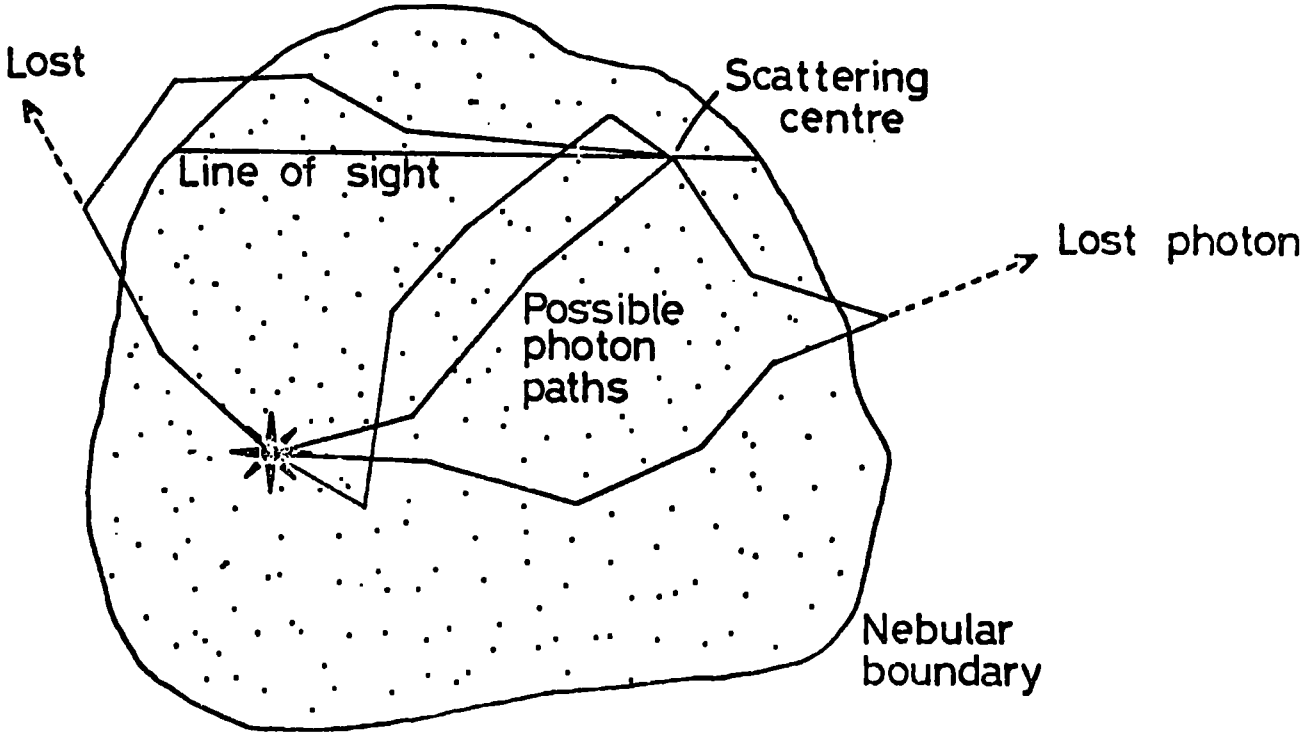


Fig.14.5 Photon loss from the nebular boundaries

$$w_G(z) = \frac{1}{4\pi r^2}\left|\frac{dr}{dz}\right| = \frac{1}{\left|4\pi\ r z\right|} \qquad 14.48$$

By this means, every scattering can be considered as if it took place on the line of sight (except for those with $r < R_o$, which lie outside the range of integration).

Consider now the introduction of finite boundaries to the nebula. In this case, the light flux on the line of sight is given not in terms of P'_n, but of some related function P''_n. Because the photons propagate independently, the boundaries affect only those which cross them, and such photons are simply lost from the nebula. Hence we can sample from P''_n simply by omitting those samples from P'_n which involve crossing the nebular boundary (Figure 14.5).

Using equation 14.19, the multiple scattering line of sight integral can be written as:

$$\underline{R} = NA\int_{z_1}^{z_2}\iint_{\underline{K},\underline{S}} t_o\sum_{n=1}^{\infty} w_G(z)\ P''_n(\underline{K},\underline{S},z)\ M(\underline{K})\ \hat{\underline{S}}\ d^3K\ d^4S \exp -\left(\frac{z_2 - z}{t_o}\right) dz \qquad 14.49$$

and hence an estimator of $\underline{R}$ is:

$$\hat{\underline{R}} = \frac{A}{mG_{EXT}}\sum_{j=1}^{m}\sum_{n=1}^{\infty} w_G(z^n_j)\ M(\underline{K}^n_j)\ \underline{S}^n_j \exp -\left(\frac{z_2 - z^n_j}{t_o}\right) \qquad 14.50$$

where $\underline{K}^n_j$, $\underline{S}^n_j$, z^n_j are samples from $P_n(\underline{K},\underline{S},z)$

14.5.8 The Final Scattering

To evaluate the above estimator, it only remains to determine the matrix M for each sample as follows:

The photon path (Figure 14.4) is transformed so that the n th scattering lies on the line of sight. In fact, due to the spherical symmetry, a number of transformations are suitable, but a rotation about the origin within the plane OEE' is used. Since it is necessary to determine if all the rotated path lies in the nebula, each scattering point is similarly transformed, as is the final reference vector for the Stokes vector, $\underline{\ell}_n$.

The final scattering is then treated as for previous scatterings by finding a new reference direction normal to the scattering plane (Section 14.5.6), the resultant Stokes vector being finally rotated into the reference frame of the observer, defined by the x - axis.

14.6 Implementation of the Monte-Carlo Scheme

14.6.1 Correlation between Samples

Because the estimator of equation 14.50 contains only sums of samples (no products), it does not matter if the samples used are themselves correlated. In fact, deliberately choosing samples which give anti-correlated contributions to the estimator is a standard technique for reducing the statistical error on the result (Hammersley and Handscomb, 1975 - antithetic variates). Consequently, we introduce no error if we derive other valid samples from each photon path used. Two samples can thus be

obtained by using both points of intersection with the line of sight column (Figure 14.4) and a degree of anti-correlation results since these samples occur in forward-backward scattering pairs. (Further efficiency may result from generating other valid samples by e.g. reflection in the x - y plane, or rotation about OE' and OF, but this has not been tried in the present context).

Another important use of this technique is that contributions from photons after (n + 1) scatterings can be found by adding a further scattering to the path used for the first n scatterings. In this way, the number of scatterings considered varies as n_T (the highest scattering order considered) instead of $\frac{1}{2}n_T(n_T+1)$.

14.6.2 Summary of the Implementation

Having described the individual steps of the technique in some detail, the following indicates the overall procedure:

i) Select a random initial direction for the photon and set the Stokes vector to (1,0,0,0).

ii) Calculate the position of the first scattering.

iii) Rotate the photon path on to the line of sight to obtain 2 points of intersection (Figure 14.4) and check to see that the rotated paths lie within the nebula - if not, do not calculate the contribution from this sample (omit step iv).

iv) Calculate the contribution to the estimator of equation 14.50.

v) Add a further scattering to the photon path and find the new Stokes vector and scattering position.

vi) Return to (iii) to find the contribution from this new photon path and repeat steps (iii), (iv) and (v) for up to n_T scatterings.

vii) Return to step (i) and repeat the process for each photon used (say $\sim$ 1000 times).

14.6.3 Numerical Details

The scattering matrix elements C,D,E,F (Equation 14.2) are determined according to equations 14.3 to 14.7 using the integration scheme developed for the functions F_1 and F_2 (Section 11.6). These elements are evaluated at 51 scattering angles and an interpolating Chebyshev polynomial expansion is then fitted to these values, enabling them to be interpolated at 201 angles corresponding to equal increments of the argument cos Θ. These results then form a table of scattering functions used as a basis when calculating the effect of a given scattering. To obtain further intermediate values, linear interpolation in $w(\Theta)C(\Theta)$, etc., is used and, since the weight function $w(\Theta)$ is specifically chosen to minimise variations in this product, the interpolation is very accurate in practice.

The "random" numbers used are all obtained by applying suitable transformations on pseudo-random samples, uniformly distributed in the range 0 to 1, obtained using

the random number generator of the NAG Fortran subroutine library.

Considerable effort has been devoted to numerical checking of the routines used, a process which is complicated by the lack of any previous work for comparison of results. The photon propagation and the polarisation at each scattering have been checked by hand calculation, and the requirement that photons are (on average) conserved provides a useful check on the calculation and normalisation of the scattering functions. In the case where only single scatterings are considered, the Monte-Carlo model reduces to the extinction only model of the previous chapter (where photons are lost on their first scattering) - hence this case has been checked against the more direct method of numerical integration.

Finally, to improve efficiency still further, the first order contribution to the estimator of equation 14.50 has been replaced by the results of the above numerical integration, resulting in a significant reduction in the statistical error, since the first order scatterings usually contribute most to the result.

14.6.4 Multiple Results and Errors

If Monte-Carlo estimates are required at a number of points in a nebula, the same set of photon paths can be used for each of the lines of sight involved. This results in a considerable saving in computer time, although the errors on the individual results are then

highly correlated. Normally, however, this does not matter.

The errors on the estimates are found by splitting the number of photons used into 10 or more batches (depending on the number). The error is then estimated from the dispersion of the results for each batch.

Because of the nature of the process, the errors on the 4 individual components of the Stokes vector are correlated to an unknown degree. Consequently the error on, say, the percentage polarisation, cannot later be easily calculated and must be estimated using the above method of evaluating it for each batch of photons used, and calculating the dispersion of the results.

14.6.5 Scattering Order

As successively higher orders of scattering are considered, the contribution to the estimator of equation 14.50 diminishes. In the presence of absorption this is usually due to the loss of light which occurs at each scattering - it is, however, also true in the absence of absorption.

The density of first order scatterings in a nebula of constant dust density is proportional to $\frac{1}{r^2} \exp(-r/t_o)$, but for subsequent scatterings this distribution spreads out, due to the random walk of photons away from the origin. Thus higher order scatterings become more important at large radial distances, becoming correspondingly less important close to the star. At any distance there is

a scattering order above which the photons are so dispersed that the contribution to the light which an observer sees is negligible.

The highest scattering order which must be considered in a given case must be determined experimentally - choosing the lowest order consistent with the required accuracy.

14.7 Application to the Model of NGC1999

14.7.1 The Qualitative Approach

Because of the relatively large computer time requirements and the stochastic nature of the results, the Monte-Carlo treatment of multiple scattering does not easily lend itself to the construction of models representing optimised fits to the data. As a result, a more qualitative approach must be adopted and the effects of multiple scattering are evaluated here by comparison with the results of the single scattering approximation at the same wavelength. To speed the computation, the single wavelength of 0.49μm has been used and, as a result, the polarisations calculated no longer represent an optimised fit to the data, even in the single scattering case. This does not invalidate the comparison with the multiple scattering results, however, since the qualitative effects seen here are unlikely to change significantly when applied to a model incorporating the full wavelength integration.

Figures 14.6 and 14.7 show the polarisation and surface brightness calculated using the best fit parameters

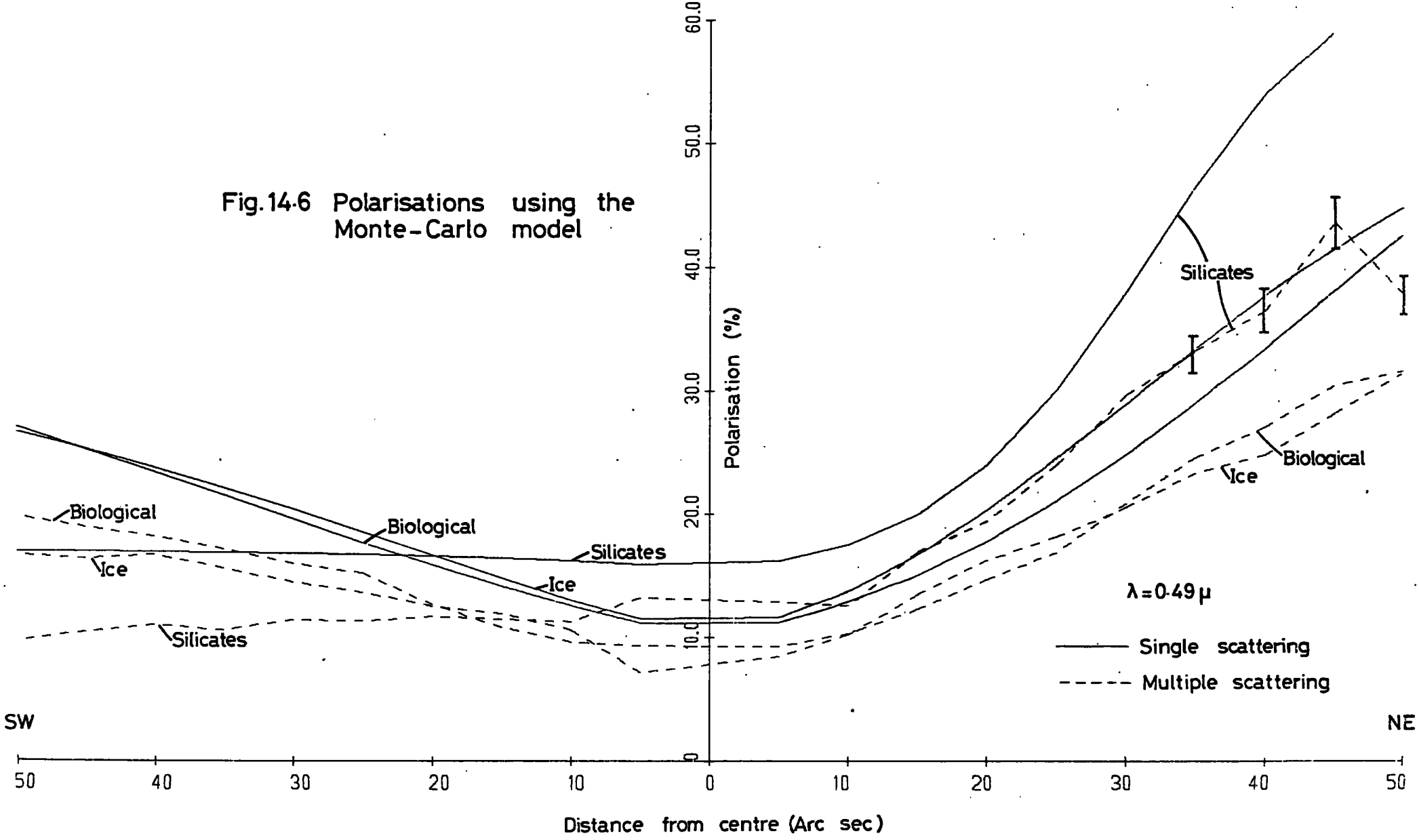
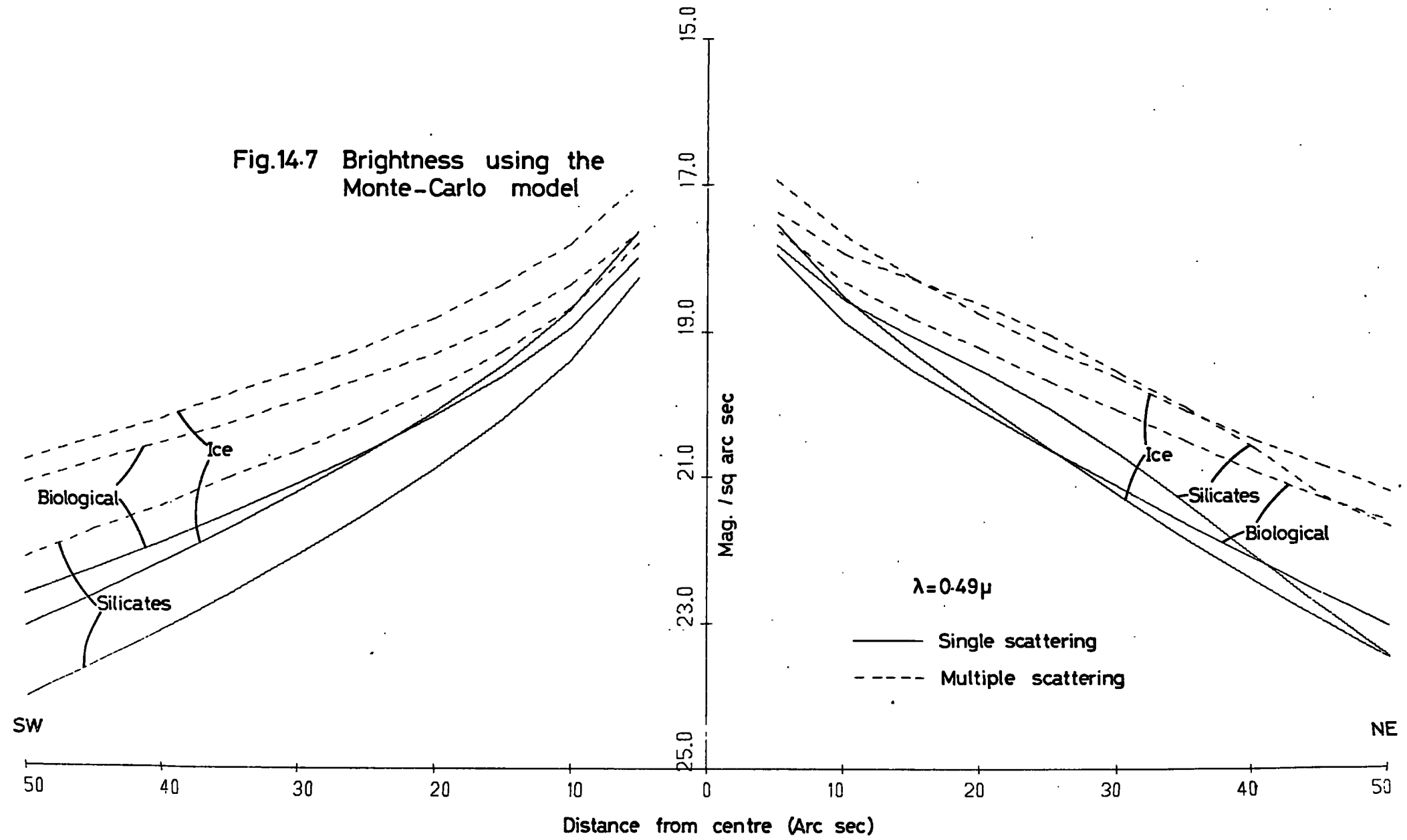


Fig. 14·6 Polarisations using the Monte-Carlo model

Fig.14.7 Brightness using the Monte-Carlo model



of table 9 (Section 13.7.2) at the single wavelength of 0.49μm. The results are shown both for single scattering (solid curves) and multiple scattering (dashed curves) and in each case the number of photons used was 2000. Unless indicated by error bars, the statistical error on the Monte-Carlo results is < 0.1 magnitude in brightness (typically 0.03) and $< 1.5\%$ in polarisation (typically 0.5%). Scatterings up to 3rd order were considered for the "biological" model and up to 4th order for the ice and silicate models.

The depolarising effect of multiple scattering is clearly seen in figure 14.6, where, for all 3 grain models the level of polarisation is reduced, although it still shows approximately the same form. The amount of depolarisation varies between factors of ~ 0.6 to ~ 0.9 and is of the same order found by Vanysek and Solc (1973) and White (1979b). The fact that very large amounts of depolarisation do <u>not</u> occur can be attributed to the forward throwing scattering functions, which mean that the scattered light is still propagating approximately radially outwards from the star, even after several scatterings. It is this effect also which causes the multiply scattered light to "disperse" rapidly from the origin as the scattering order increases, so that the highest order of scattering which need be considered is much lower than if the scattering functions were isotropic.

The effect of multiple scattering on the calculated surface brightness (Figure 14.7) is to increase the overall

level of brightness, while decreasing the brightness gradient - a result expected from the increasing importance of the multiply scattered light at large offset distances. The effect on the surface brightness is probably more significant than the relatively small effect on the polarisation, since it amounts to almost 2 magnitudes at 50″ from the star and the overall form of the brightness is also changed - in particular the asymmetry seen in the single scattering results is almost completely removed when multiple scattering is included.

14.7.2 Adjustments to the Model

The main consequence of the large effect which multiple scattering has on the surface brightness, compared with the small effect on polarisation, is that, while the scattering properties of the dust grains will not have been greatly mis-calculated by using the single scattering approximation, their number density will have been considerably over-estimated. Consequently, some further adjustment of the dust density is required in order to return the surface brightness to the level given by the single scattering approximation. (Which provides a good fit to the data at an offset distance of 20″.) The following table shows the dust densities required to achieve this:

Grain Type	N (m^{-3})
Biological	0.016
Ice	0.014
Silicates	0.021

Table 11

The surface brightness resulting from this reduction in dust density is shown in figure 14.8 (together with the original single scattering results using the higher density). With the reduced dust density, scatterings up to 3rd order provided sufficient accuracy for all 3 grain models and the statistical errors are similar to the calculations with the higher density. It can be seen that the lower brightness gradient and more symmetric distribution of brightness, caused by the multiply scattered light, are still present and a glance at figure 13.14 shows that these are precisely the features required to match the observed surface brightness in NGC1999.

As expected, the reduced amount of multiple scattering, resulting from a reduction in dust density, means that the depolarisation produced is somewhat less than before and the results are shown in figure 14.9. This small amount of depolarisation, coupled with the very rapid change of polarisation resulting from a change in the mean grain size (Figures 11.6 to 11.10, 12.6 and 12.7), means that the fits to the polarisation data could probably

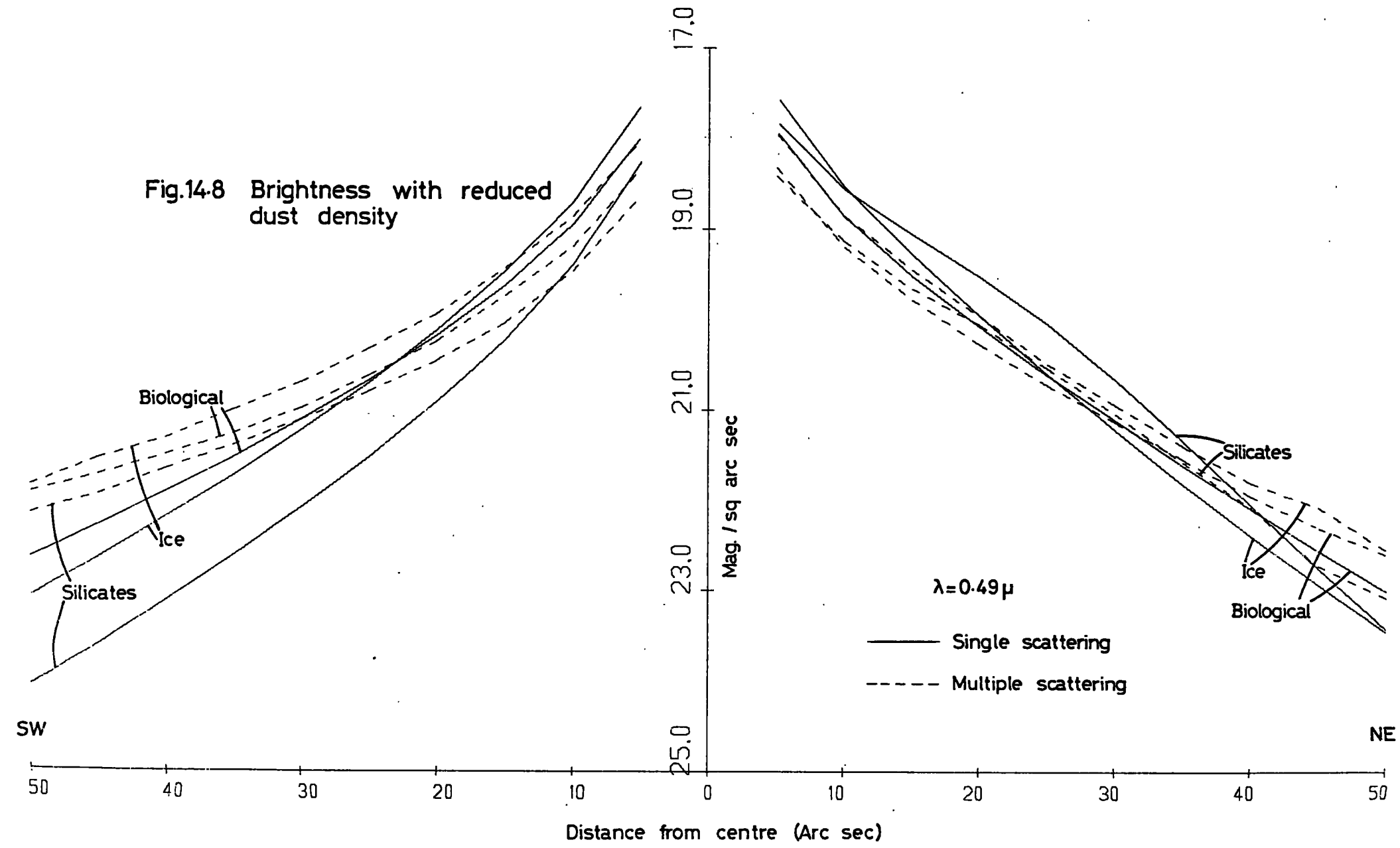
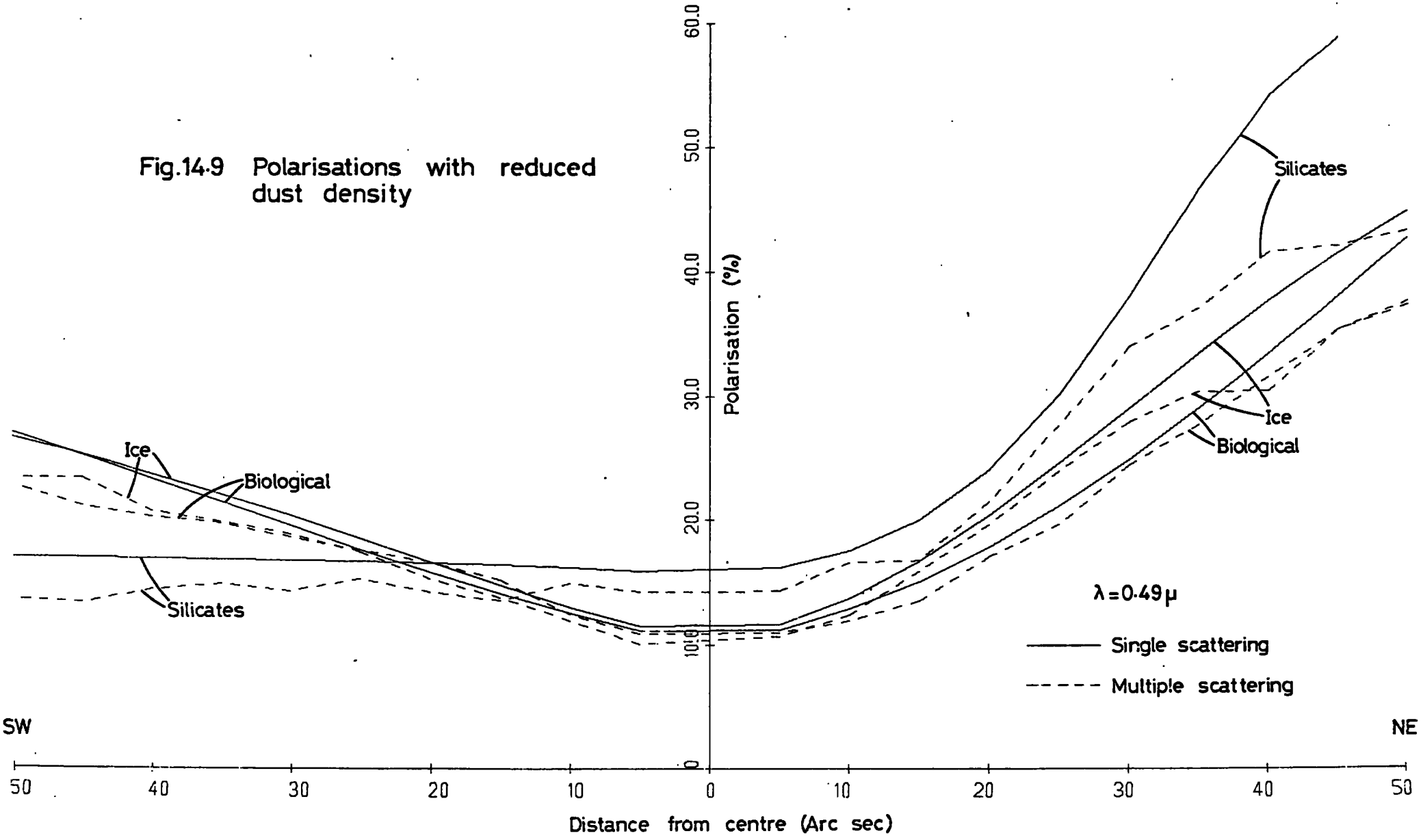


Fig.14·8 Brightness with reduced dust density

Fig.14·9 Polarisations with reduced dust density



be restored in the presence of multiple scattering by a reduction of the mean grain size by at most a few percent.

14.8 Conclusions

Application of the new Monte-Carlo technique to the model of NGC1999 derived in the previous chapter shows that, by using a single scattering approximation which includes internal extinction, the surface brightness is considerably underestimated and the resulting grain density is too high. When correction is made for this, the calculated surface brightness appears able to explain the observations using any of the biological, ice or silicates grain models and no significant brightness asymmetry is caused by the tilt introduced to explain the polarisation data. Using the revised dust densities of table 11, the resulting stellar extinctions predicted by the models at a wavelength of 0.5μm are as follows:

Grain Type	Stellar Extinction (mag.)
Biological	0.72
Ice	0.83
Silicates	1.07

Table 12

and all of these values are probably consistent with the observed stellar reddening (Section 10.2.3).

In contrast to the brightness, calculation of the depolarisation caused by multiple scattering shows that in NGC1999 the effect is quite small, and does not greatly affect the grain size parameters derived using the single scattering approximation. These calculations thus support the conclusion of White (1979b) that the observation of high values of polarisation does not necessarily indicate the absence of multiple scattering, with its associated effect on the surface brightness and colour of the nebula. As a result, polarisation data are likely to be a far more reliable indication of nebular parameters (particularly geometry and grain size) than brightness and colour data. This relatively high information content of polarisation data, compared with intensity data, in the presence of multiple scattering has also been commented on by Coffeen and Hansen (1974) in the context of planetary atmosphere studies.

It is important to note that NGC1999 is a particularly bright nebula relative to its illuminating star - as judged by the fact that the stellar profile is not seen distinctly from the inner nebular contours - and the effects of multiple scattering are thus likely to be particularly important. As a result, nebulae which are relatively more faint are likely to have a very small depolarisation, and polarisation data for these nebulae can probably be interpreted reliably in the single scattering approximation - such may not be true of their brightness, however. A

further discussion of ways in which the single scattering approximation may be made more accurate will be given in the next chapter.

CHAPTER 15

DISCUSSION

15.1 Introduction

In the previous two chapters, pure reflection nebula models of NGC1999 have been developed using a number of refractive indices for the nebular dust. These models have been judged on their ability to explain both the polarisation and brightness data currently available and these criteria were found to provide sensitive tests of nebular models - in particular, the mean grain size and dust density appropriate to a given refractive index are well defined.

In this chapter, these models will be examined in a wider context and the relevance of the resulting optical properties of the nebular dust to the nebular colour, measurements in other reflection nebulae and to the observed interstellar extinction law will be investigated. Studies of interstellar polarisation, however, cannot be directly compared in this context, since this phenomenon depends on the presence of non-spherical dust grains, the effects of which have been neglected in NGC1999.

15.2 Nebular Colour

As discussed at the end of chapter 14, the effects of multiple scattering within a nebula have a far greater effect on the calculated surface brightness than on the polarisation. Thus the calculation of nebular colours, being the ratio of the surface brightness at two wavelengths, is particularly sensitive to finite optical depth

effects and, for this reason, discussion of the colour of NGC1999 has been left until now, when complete models are available which include multiple scattering effects.

Nebular (B-V) colours have been calculated using the best-fit model parameters of table 9 (13.7.2) and the revised dust densities of table 11 (14.7.2). Reddening of the illuminating star by extinction within the nebula has been found from the extinction predicted by the models (Table 12) using a ratio of total to selective extinction, R = 3.3. This gives reddenings of typically 1 or 2 tenths of a magnitude - consistent with the observations (10.2.3). The full multiple scattering model has been used with 2000 photons at single B and V wavelengths and the resulting (B-V) colours are shown in figure 15.1. Typical statistical errors on these results are $\pm$0.05 mag.

All three grain models considered give nebular colours which are very similar, and very accurate colour measurements would clearly be needed to distinguish between them. The predicted nebular colours all vary far more slowly with offset distance than is the case if multiple scattering is neglected, and little asymmetry in nebular colour results from the asymmetric geometry - although, as with surface brightness, most asymmetry is seen in the silicates model, probably on account of its large nebular tilt. All three models give nebular colours which must be considered consistent with the limits set by the photographic study of Brück (1974) (Dashed lines in Figure 15.1).

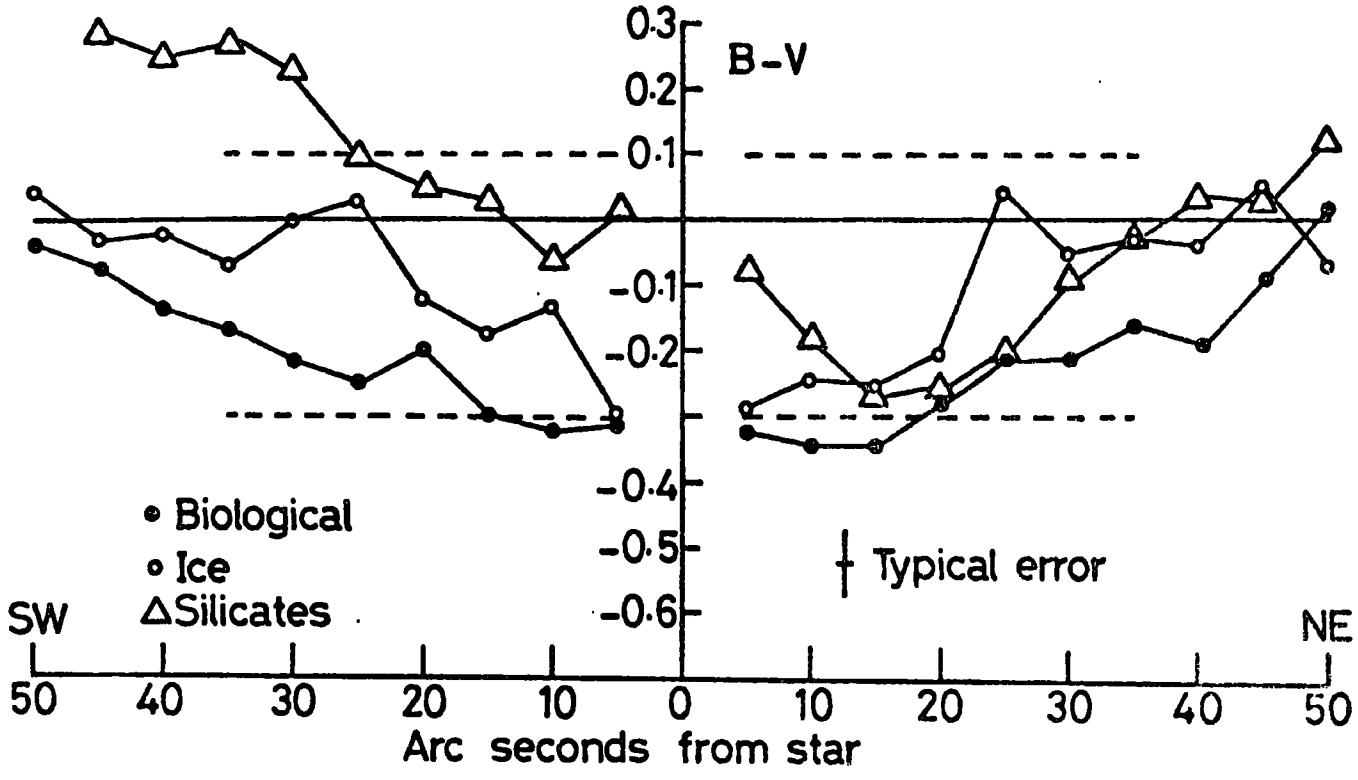


Fig.15.1 Monte-Carlo colour predictions

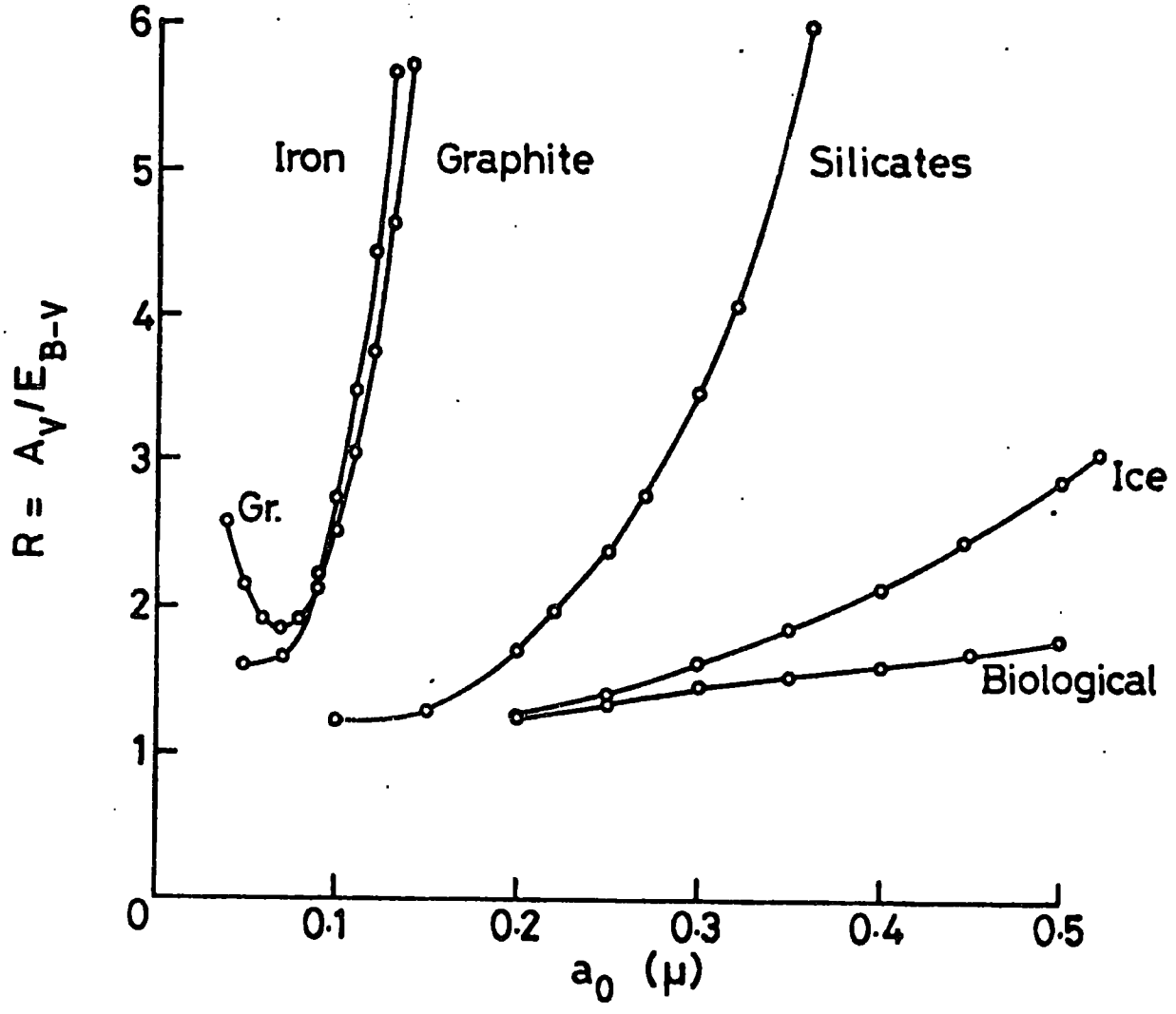


Fig.15.2 Reddening factors

15.3 Conclusions about Nebular Models

15.3.1 The Model of NGC1999

The studies of chapter 13 and 14 have indicated that, even in the presence of moderate amounts of multiple scattering, polarisation data can be used as a good indication of certain nebular parameters and that, in conjunction with brightness data, models can be constructed which give well defined mean grain sizes, dust densities and geometries if the refractive index and size distribution of the dust grains is assumed.

Summarising the results obtained for various grain models, it can be said that in NGC1999 highly absorbing grains with low albedos, such as iron and graphite, introduce too much extinction to explain either the observed nebular brightness or the comparatively small stellar extinction. These results also have difficulty explaining the observed range of polarisation - a deficiency reflected in their small polarising range, η , which is probably related to scattering functions which are insufficiently forward throwing (12.6).

Models based on grains with a higher albedo, however, appear capable of explaining the nebular brightness, the stellar extinction and the nebular colour; the asymmetry introduced to explain the polarisation observations having little effect on the brightness and colour predictions. Complete multiple scattering calculations for the three cases considered (with refractive indices 1.16, 1.33 and 1.63 - 0.05i) show that no distinction can be drawn on

the basis of brightness or colour data. The polarisation data, however, restricts the choice of model somewhat more and, while the two lowest refractive indices considered (biological and ice) give fits to the polarisation data which may be considered satisfactory (although small perturbations to the geometry may be required to explain the data in full), the model based on the silicates refractive index ($m = 1.63 - 0.05i$) has some difficulty in explaining the observed range in polarisation - a result again reflected in the small polarising range, η, for these grains. As discussed in section 12.6, this conclusion is virtually unaffected by the choice of nebular geometry, so long as the illuminating star is contained within the nebula - a situation which seems unavoidable in NGC1999 due to the high surface brightness (10.4). The significance of this conclusion may be affected by the choice of size distribution function, however. (See section 15.6.)

15.3.2 The Single Scattering Approximation

In addition to specific conclusions drawn about NGC1999, this study has indicated that important approximations may be made in the interpretation of nebular data. Calculation of the effects of multiple scattering (14.7) show that, even for nebulae such as NGC1999 where finite optical depth effects are likely to have above-average importance, the amount of depolarisation expected relative to a single scattering model is rather small.

The explanation of this effect lies in the very forward throwing scattering functions which result from the use of the grain sizes necessary to explain the observed levels of polarisation. Most photons in a nebula thus undergo only small angle scatterings (which introduce little polarisation) and can propagate in approximately straight lines for distances many times the classical extinction length within the nebula. In this situation it is clear that if the grain albedo is high the use of a model which includes internal extinction, and thus assumes all photons to be lost at their first scattering, will introduce considerable error in the calculation of nebular brightness. Indeed, complete multiple scattering calculations give brightness predictions which agree more closely with the simplest model which neglects extinction, then with the model which includes extinction in the manner used by e.g. Hanner (1971).

Davidson and Ruiz (1975), in constructing a scattering model of Eta Carinae, have followed the example of Werner and Salpeter (1969) and noted that if the scattering function is forward throwing the extinction cross section may be better approximated (for the purposes of scattering calculations) by $(1-g)\ G_{SCA}$, where g is the phase parameter of the grains (the mean value of the cosine of the scattering angle) and has a value close to 1 for forward throwing grains. Extending this idea to absorbing grains, the use of an effective cross section

$$G_{EFF} = G_{ABS} + (1 - g)\, G_{SCA} \qquad 15.1$$

might be appropriate in calculating the brightness and polarisation of reflection nebulae in the single scattering approximation. If such an approximation is used, it seems likely that polarisation data in nebulae with optical depths lower than in NGC1999 may be reliably interpreted in terms of single scattering models, and the associated brightness determinations might also be sufficiently accurate at small optical depths.

15.4 Comparison with Interstellar Extinction Data

15.4.1 Reddening Factors

The mean interstellar extinction law in the wavelength region 0.9μm to 0.33μm is described to a good approximation by an optical depth which increases linearly with $1/\lambda$. (For a review of extinction observations in this wavelength range, see Wickramasinghe and Nandy, 1972). A parameter commonly used to describe the observed extinction law is the ratio of total to selective extinction, R, in the blue-visual region of the spectrum:

$$R = \frac{A_V}{E_{B-V}} = \frac{C(\lambda v)}{C(\lambda_B) - C(\lambda_V)} \qquad 15.2$$

where $C(\lambda)$ is the wavelength dependent cross section of the interstellar grains which cause the extinction. If

the optical depth for extinction were strictly proportional to $1/\lambda$, the resulting value of R would be 4. In fact, slight deviations from this law give a rather smaller value and $R \sim 3$ appears to be a generally accepted average value for many regions of the galaxy, although local variations are seen.

Using the grain size distribution of equation 11.17 the ratio R is simply calculated from equations 11.14 and 15.2 and is shown in figure 15.2 as a function of the size parameter a_o for various refractive indices. It is on the basis of curves such as these that R is usually taken to be an indicator of the grain size responsible for the extinction.

The values of R resulting from the grain sizes required in the model of NGC1999 are summarised below. Also shown are the values of a_o which give R = 3 for each refractive index.

Refractive Index	a_o (μm)	R	a_o (μm) for R=3
1.16	0.324	1.49	>0.9
1.33	0.265	1.48	0.51
1.63-0.05i	0.174	1.46	0.28

Table 13

It can be seen that the polarisation data has selected grain models which not only have similar polarising properties,

but also very similar reddening laws. The grain sizes, however, are significantly smaller than those appropriate to the interstellar extinction law and give an amount of reddening in the blue-visual region which differs substantially from that commonly observed.

15.4.2 Relevance of Nebular Observations

In view of the discrepancy noted above, it is important to ask to what extent the results of nebular observations should be compared with the interstellar extinction data - in particular, is the dust which populates nebulae such as NGC1999 the same dust as is responsible for the extinction law?

Perhaps the most notable feature of the extinction law is that the optical properties of the interstellar dust appear to be remarkably constant throughout most regions of the galaxy. Consequently, it appears likely that many observable reflection nebulae should arise as a result of illumination of dense regions of this dust by nearby or embedded stars and, if this is the case, a similar constancy of optical properties should be seen in reflection nebulae.

The early work of Hubble which lead to an empirical relation between the observable radius of a reflection nebula and the brightness of the illuminating star (Equation 10.1) suggests that some constancy of optical properties is indeed observed in many reflection nebulae.

Polarimetric observations of a number of reflection nebulae made with the Durham polarimeter have indicated that the levels of polarisation found in many nebulae are remarkably similar. The maximum level of polarisation, in particular, often takes values between 30 and 40%. This finding is doubly surprising, since not only are geometric factors expected to influence the level of polarisation observed, but polarisation is also a very sensitive indication of the size of dust grains (Figures 11.6 to 11.10, 12.6 and 12.7) - probably more sensitive, in fact, than the reddening factor R (Figure 15.2). Some of these observations are summarised briefly below:

Nebula	$\lambda(\mu m)$	Max Polarisation (%)	References
NGC1999	0.49	30 to 40	This work (10.3)
NGC 6726/7	0.55	30 to 40	This work (9.6)
NGC 6729	0.55	30 to 40	Warren-Smith and Scarrott (unpublished)
Eta Carinae	0.5	~35	Warren-Smith et Al. (1979)
Lk $H\alpha$ 208	0.5	~30	Scarrott, Warren-Smith and Watts (unpublished)
H - H 101	0.55	~20	Scarrott and Warren-Smith (unpublished)
M43	0.55	~25 increasing at edge of field.	Khalesse et Al. (1979)

Table 14

Presumably, at large offset distances from the illuminating stars, (where the maximum polarisation is always observed) the geometries in most reflection nebulae become asymptotically similar, in the manner of figure 12.5. Clearly, however, these observations require that the size, and possibly nature, of the dust grains which produce this polarisation must be very similar in all these nebulae, even although the nebulae themselves differ greatly in type and probable origin.

NGC1999 appears to be a typical nebula, in that it satisfies the Hubble relation (Equation 10.1) and has polarisations typical of other nebulae.

In addition to the above, there is some evidence that a typical interstellar extinction law must apply in NGC1999, since use of the stronger reddening laws of table 13 results in stellar reddening somewhat greater than observed and nebular colours approximately 0.5 magnitudes bluer, which are inconsistent with the observations of Brück (1974). Use of the classical extinction law, however, gives colours in agreement with observation.

Although the model developed to account for the observations of NGC1999 appears able to explain the polarisation, surface brightness and colour data, the possibility remains that, perhaps due to the use of an inappropriate geometry, the grain sizes resulting from the model are in error. It is not difficult to see, however, that the grain sizes necessary to explain the extinction law (table 13, column 4) produce polarisations

which are too small to explain the observed levels of 30 to 40% with virtually any nebular geometry. Consequently, the conclusion remains that while it appears likely that most reflection nebulae are populated by interstellar dust for which a typical interstellar extinction law is appropriate, it does not appear possible using the size distribution of equation 11.17 to construct a grain model which can simultaneously explain both the nebular polarisations and the extinction law. Both sets of observations may be explained separately, however, by appropriate adjustment of a_o.

15.5 Implications for Grain Models

15.5.1 Difficulties with Present Models

Following the discussions above, the following summarises the problems associated with the explanation of nebular polarisations:

i) Typical nebular polarisations require grain sizes which are significantly smaller than those needed to explain the interstellar extinction law. It seems likely, however, that the same dust is responsible for both phenomena.

ii) Using the size distribution $n(a) \propto \exp\left[-5(a/a_o)^3\right]$, refractive indices appropriate to silicates (and larger indices) result in grain models with a polarising range, η, which is too small to explain the typical large range of polarisations observed. Thus the construction of nebular models using silicate grains is very difficult, although

they appear to be a likely component of interstellar dust and the 10μm spectral feature, commonly attributed to silicate material, is often observed. The small polarising range is probably associated with scattering functions which are insufficiently forward throwing.

15.5.2 Possible Solutions

A number of ways of overcoming these problems may be possible.

Voshchinnikov (1978) has investigated the effect of inhomogeneous grain size distributions - in particular, depletion of small grains close to the illuminating star as a result of expulsion by radiation pressure. His calculations show an expected lowering of polarisation levels close to the star relative to the no-depletion model and by this means it may be possible to construct models with a range of polarisation which exceeds the value of η for the grains used. A quantitative description of this effect is, however, difficult, since the depletion depends on rather uncertain quantities such as the luminosity history of the star and the terminal velocities of grains in the nebular medium. This approach does not solve the problem of small grain size, since it is likely to result in an increase in mean size by the removal of the smaller grains.

It is likely, of course, that interstellar dust may contain more than one component, with differing grain sizes and refractive indices. It may thus be possible to

construct a grain mixture in which one component is dominant in extinction, but, being highly absorbing, scatters little light so that the scattering properties are dominated by another component with a different grain size. Graphite has often been proposed as an absorbing component of interstellar dust, since the absorption feature in the observed extinction law at $\lambda \sim 0.21$ m corresponds to an absorption in the graphite spectrum. Good fits to the extinction law can be obtained using grain mixtures which contain graphite (e.g. Mathis et Al., 1977) but the narrowness of the absorption feature requires that the graphite grains be rather small. At first sight, therefore, it appears that the aborbing component of a grain mixture must have the smaller grains - contrary to what is required. However, the 3 component mixture of Wickramasinghe and Nandy (1972) is known to have highly polarising properties (See section 13.5.1), although it does not provide nearly so good a fit to the extinction law as that of Mathis et Al. Clearly, further numerical studies are required to see if the polarising properties can impose useful constraints on the choice of grain-mixture models of the interstellar dust.

The use of Core-Mantle grains, consisting of a refractory core surrounded by an accreted mantle of ices may also provide a solution to the problem of nebular polarisations. Two mechanisms appear possible: either the mantles are lost due to the increase in grain temperature in the vicinity of the illuminating star of a reflection

nebula, resulting in smaller sized grains (Although this does not solve the problem of the small polarising range characteristic of the refractive indices of refractory materials), or it is possible that the scattering properties of core-mantle grains may be primarily determined by their mantles, while the extinction properties may be determined by their cores (or vice-versa, as required). Numerical work is again required to determine whether this latter explanation is feasible.

The effects of grain non-sphericity may also have an important effect on the effective scattering size of dust grains if, for example, the extinction properties depend primarily on the larger grain dimension, while the reflection properties depend on the smaller dimension. The work of Greenberg and Hong (1973) and Rogers and Martin (1979), however, suggests that the extinction properties are primarily determined by the smaller dimension. It is known, also, that a randomly orientated collection of very elongated, small particles (Rayleigh needles) produce depolarisation relative to non-elongated Rayleigh particles and hence appear larger - a result attributable to the fact that non-spherical particles scatter most strongly when the electric vector is parallel to their longest dimension. Clearly this effect will operate in the wrong direction, making non-spherical particles appear larger in reflection than in extinction.

The relevance of the adopted grain size distribution to the problems of nebular polarisations has so far been ignored and, throughout this study, a function of the form:

$$n(a) \propto \exp\left[-5\left(\frac{a}{a_0}\right)^3\right] \qquad 15.3$$

has been used. As was previously noted, the advantage of this form is that the distribution of particle sizes which contribute to the scattered light is fairly narrow and symmetrical (Figure 11.4) and the mean scattering size $\langle a \rangle_{SCA}$ (Equation 11.21) is well defined. This size function thus performs its simple task of smoothing the otherwise very irregular scattering functions of single spheres by including the contributions from neighbouring sizes. As a result, models based on the use of this function give a well defined effective scattering size, but no information about the range of sizes contributing to the scattered light. It is interesting to investigate the effect of a broader size distribution function, particularly since imaging polarimetry may be able to provide an observational test of size distribution functions.

15.6 The Size Distribution Function

15.6.1 Effect on the Grain Properties

Hong and Greenberg (1978) have noted that, for a given size distribution, the mean grain size averaged over extinction efficiency differs slightly from that averaged over polarising efficiency (for polarisation by extinction), and they use this as a test of size distribution functions. In view of the small difference in sizes, however, and

the simplicity of the polarisation model (perfectly aligned, spinning, infinite cylinders) it may be more accurate to apply this test to the case of polarisation by scattering where the size difference is large and the model (at least as regards the polarising properties of the grains) is more certain. In this case it is easy to see that, since large grains have scattering functions which are more forward throwing than for small grains, they will make a relatively larger contribution to the total extinction cross section, than to the scattering cross section for angles of $\sim 90^{\circ}$ typically found in reflection nebulae. Thus, if a sufficiently wide range of grain sizes is present, the mean effective size for extinction:

$$\langle a \rangle_{EXT} = \frac{\int_0^{\infty} a\, n(a)\, C_{EXT}(a)\, da}{\int_0^{\infty} n(a)\, C_{EXT}(a)\, da} \qquad 15.4$$

may be considerably larger than that for scattering through 90°:

$$\langle a \rangle_{90^{\circ}} = \frac{\int_0^{\infty} a\, n(a) \left[|s_1(90^{\circ})|^2 + |s_2(90^{\circ})|^2 \right] da}{\int_0^{\infty} n(a) \left[|s_1(90^{\circ})|^2 + |s_2(90^{\circ})|^2 \right] da} \qquad 15.5$$

The inclusion of large grain sizes in this way also solves the problem of the limited polarising range of some grain models, for the inclusion of large grains which are forward throwing has the largest effect in reflection nebulae at small offset distances, where the range of scattering angles along the line of sight includes small-angle forward scatterings. Thus depolarisation occurs at small offset distances, as required to explain many nebular observations. Calculations show that this effect is reflected in an increased polarising range, η, for grain models as the width of the size distribution function is increased and the scattering functions become more forward throwing.

At present, it is not known how broad a size distribution function is required to explain both the nebular polarisation observations and the interstellar extinction law. Work by Mathis et Al. (1977), however, is particularly interesting in this context.

These workers performed an extensive analysis of grain mixture models appropriate to fit the interstellar extinction law. They considered combinations of up to 3 grain materials with size distribution functions decomposed into steps, with a large number of degrees of freedom, and thus able to assume many functional forms. They also took atomic abundance data into account in placing constraints on their models. They concluded that good fits to the most recent extinction data in the

wavelength range 0.11μm to 1.0μm could be obtained using graphite grains (to explain the λ = 0.21μm feature) in conjunction with almost <u>any</u> other grain material. Their size distributions, optimised to fit the data, took on forms approximating to power laws in nearly all cases, i.e.:

$$n(a) \propto a^{-\gamma} \qquad 15.6$$

with values of the index γ in the range 3.3 to 3.6.

It is simple to show algebraically that, for a constant refractive index, a power law size distribution leads to an extinction cross section which varies like:

$$G_{EXT}(\lambda) \propto \lambda^{3-\gamma} \qquad 15.7$$

Hence $\gamma \sim 4$ yields an extinction law which approximates to that observed in the wavelength range $\lambda \gtrsim 0.3$ m, for <u>any</u> refractive index. In the light of this result, the conclusions of Mathis et Al. are easily understood (they also allow for wavelength dependent optical properties). A size distribution of this form must, in practice, have a cut-off at small sizes, in order that the total grain volume be finite. The effects of this should be most apparent in the ultraviolet, and the distributions of Mathis et Al., which fit the UV data, predict such a cut-off.

It is of interest to determine whether such a size distribution function can explain the nebular polarisation data. Preliminary calculations show that to obtain the

same mean scattering sizes, $\langle a \rangle_{90^o}$, as are obtained when modelling NGC1999 (See table 15), values of γ from 4.5 to 5 are required. (Simple arguments show that $\gamma = 4.3$ gives $R = 3.3$ for a constant refractive index). This, however, is only an approximate comparison, since use of a different scattering function resulting from a greatly different size distribution may alter the geometric parameters required to represent NGC1999 and thus alter the value of $\langle a \rangle_{90^o}$ required. Although this change is likely to be small, further modelling of the data with a power law size distribution is needed for a full investigation.

15.6.2 Spectral Data

As was noted in section 11.8, the scattering functions appropriate to a given distribution of grain sizes are to a good approximation given by a function of the mean scattering size $\langle a \rangle_{SCA}$ (Equation 11.21). If the wavelength of observation is varied it is the variation of the dimensionless parameter:

$$x_{SCA} = \frac{2\pi \langle a \rangle_{SCA}}{\lambda} \qquad 15.8$$

which primarily determines the wavelength variation of the scattering functions and hence of the observed polarisation. In interpreting spectral data, therefore, it is important to allow for the variation of $\langle a \rangle_{SCA}$

with the wavelength of observation.

Differentiation of equation 15.8 shows that an observable, C, which depends on the Mie scattering functions, and hence on the dimensionless parameter X appropriate to the effective mean size of the grains $\langle a \rangle$, has a wavelength dependence described by:

$$\frac{dC}{d\lambda} = \frac{dC}{dX} \times \left[\frac{d}{d\lambda} \ln \langle a \rangle - \frac{1}{\lambda} \right] \qquad 15.9$$

The first 2 terms in this expression are size and model dependent and can thus provide a measure of the effective mean size (or of X) from observation of the spectral variation of C. This has frequently been done for the interstellar extinction data. The last term, however, depends on the size distribution function and information about this term cannot be obtained from spectral observations of C. Clearly, then, size estimates based solely on interpretation of spectral data may be considerably in error if the size distribution assumed differs greatly from the true distribution - indeed, it may be true (if the final term above dominates the spectral variation of C) that observations of C provide more information about the size distribution than about the effective mean size and properties of the dust grains (e.g. Mathis et Al., 1977).

The variation of $\langle a \rangle$ with wavelength, which determines the value of the distribution-dependent term above, depends

on the balance between the increasing grain cross section ($\sim a^2$) and the functional form of the size distribution cut-off at large grain sizes and on the way in which this balance changes with wavelength. In the limiting case of a single grain size $\langle a \rangle$ is constant, and the spectral variation of polarisation is maximised. Conversely, it is easy to show that a power-law size distribution (Equation 15.6) results in an effective mean grain size which varies in proportion to the wavelength of observation and hence gives wavelength-independent polarising properties for the collection of dust grains. Comparison of the measured wavelength dependence of polarisation in nebulae with that expected for a constant effective mean size can thus yield information about the wavelength dependence of $\langle a \rangle$ and hence about the functional strength of the size distribution cut-off. In general, strong wavelength dependence indicates a strong cut-off, while weak wavelength dependence indicates a weak cut-off and consequently (Section 11.8) a contribution to the scattering functions from a wide range of grain sizes.

The number of published spectral observations of polarisation in reflection nebulae is very large and a survey will not be attempted here. Zellner (1974) gives a review of some of this work and gives examples for a few well studied nebulae. The work of Serkowski (1969) on NGC6726/7 is also relevant to the observations of

section 9.6. Inspection of figures 11.6 to 11.10, 12.6 and 12.7 show that for the size distribution used in this study (for which $\langle a \rangle$ is by no means constant - see figure 15.3) the typical variation of polarisation with wavelength (equivalent to varying the parameter a_0 in these diagrams) is extremely rapid. In all cases considered, the full range of polarisation available, from zero to saturation at the maximum level, is encountered when the wavelength is varied by a factor of from 2 to 3. This fact is also reflected in the model of NGC1999 (Chapter 13) where knowledge of the overall level of polarisation and the wavelength gives an extremely well-defined grain size for a given model geometry.

The typical variations found in real nebulae, while sufficiently large to be easily measurable, are weak in comparion with those expected and, although individual workers have managed to explain some of the available data in terms of specific scattering angles, grain sizes and refractive indices, the adoption of a grain size distribution with a relatively weak cut-off can clearly provide a general explanation for a large number of these observations. It is important to note, also, that in many nebulae, the depolarisation resulting from finite optical depth effects can probably explain much of the polarisation variation observed, since these effects vary very rapidly with wavelength due to the form of the interstellar extinction law. Clearly the construction of nebular models is necessary, in cases

where finite optical depth is present, before spectral polarisation data can be precisely interpreted.

The importance of imaging polarimetry in this context lies in the fact that the observed level of polarisation can be related, via nebular models (such as has been described in this thesis for NGC1999), to a well-defined estimate of the effective mean grain size, $\langle a \rangle_{90^o}$, in the nebula, <u>without</u> the need to vary the wavelength of observation. Estimates of mean grain size derived from polarimetry at a single wavelength are thus largely independent of assumptions about the size distribution function and can be used to investigate the wavelength dependence of $\langle a \rangle_{90^o}$ and hence to provide a test of size distribution functions. The effective mean scattering sizes derived from the model of NGC1999 by evaluation of equation 15.5 are shown below for the values of a_o in table 9 ($\lambda = 0.49\mu m$):

Refractive Index	a_o (μm)	$\langle a \rangle_{90^o}$ (μm)
1.16	0.324	0.169
1.33	0.265	0.150
1.63-0.05i	0.174	0.113

Table 15

Figure 15.3 shows the expected variation in the mean effective scattering size (normalised to $\lambda = 0.49\mu m$) for a single grain size, a power-law size distribution

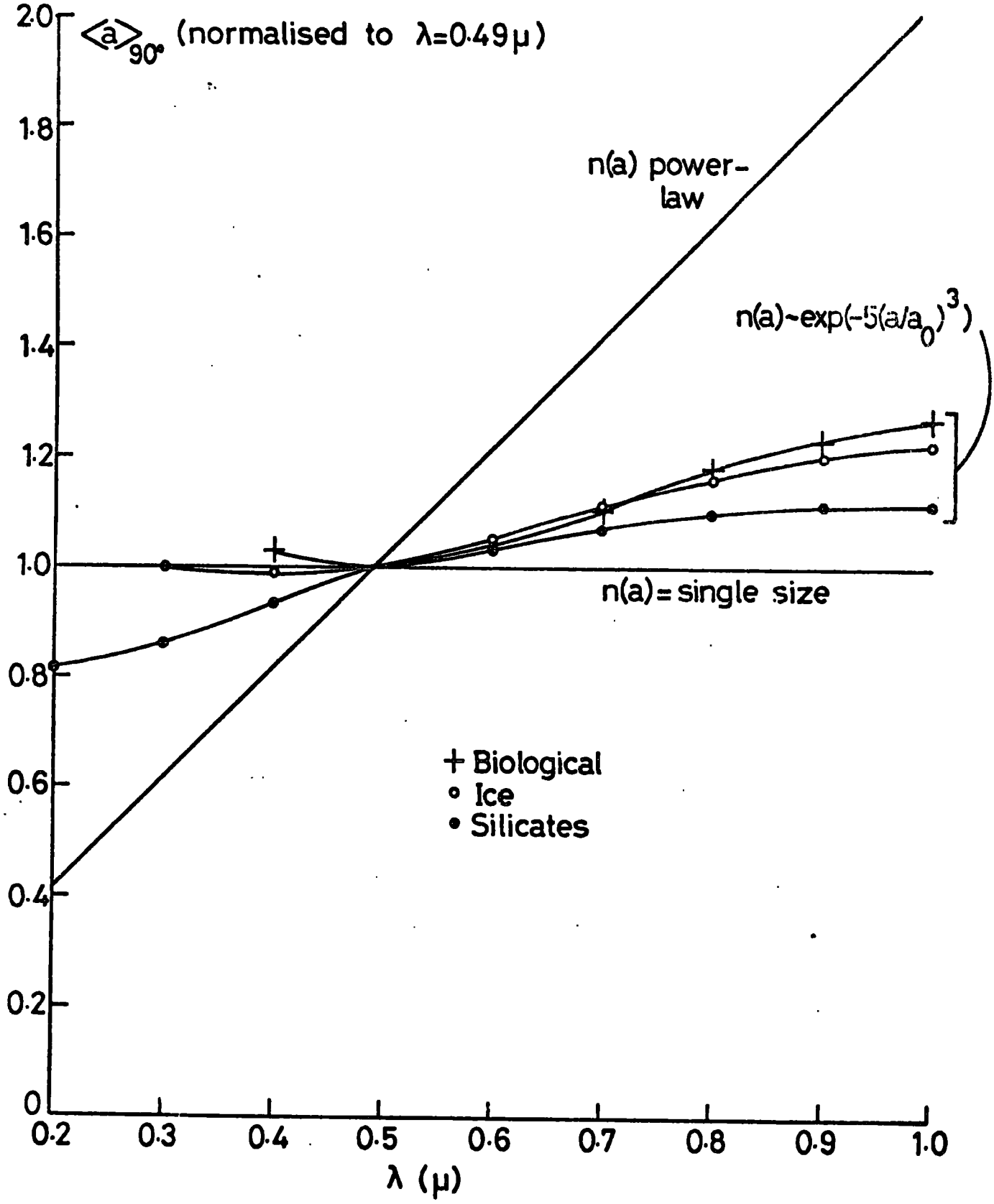


Fig.15.3 Apparent size of scattering grains as a function of wavelength

and the form used in the study of NGC1999 (Equation 15.3). The results expected can vary widely depending on the choice of size distribution and, at least for longer wavelengths, the effect of the choice of refractive index is not large. It should certainly be possible, by constructing nebular models at other wavelengths, to discriminate between "quasi-power-law" distributions and those in which $\langle a \rangle$ is relatively static.

Evaluation of equivalent curves for the size distributions of equations 11.18 and 11.19 give results which are virtually indistinguishable from the results for $n(a) \sim \exp\left[-5\ (a/a_o)^3\right]$, and clearly the range of size distributions usually considered represents (at least as far as wavelength dependence of scattering properties are concerned) only a small subset of that possible.

15.7 Conclusions

The problems arising in the interpretation of typical nebular polarisations (15.5.1) appear to have 3 classes of possible solution:

i) The dust which populates many reflection nebulae is not typical of the dust which produces the interstellar extinction, although it appears to have constant properties in most nebulae, and the wide range of polarisation found in some nebulae results from depletion of small grains close to the illuminating star.

This solution suggests that illuminating stars have a large physical effect on the surrounding dust.

ii) The interstellar dust contains several distinct components, either separate or in the form of core-mantle grains, in an unknown combination, such that the component which dominates the extinction properties is different from that which dominates the polarising properties for large angle scatterings.

iii) The size distribution function adopted is too narrow, and a broader form, possibly approximating to a power-law is appropriate for the interstellar dust.

This latter solution appears particularly attractive as it can simultaneously explain many features of the interstellar dust:

i) The work of Mathis et Al. (1977) indicates that a power-law size distribution can provide excellent fits to the mean interstellar extinction data over a wide wavelength range, and that the fit is virtually independent of the choice of refractive index, so long as graphite is included as an extra component to fit the 0.21μm extinction feature. These results indicate that for such a size distribution the extinction properties at visual wavelengths are relatively insensitive to changes in refractive index, such as might result from composition variations or the accretion of icy mantles by interstellar grains. These conclusions are supported by simple theoretical calculations.

ii) Simple calculations show that the levels of polarisation in NGC1999, which are typical of many nebulae, may be explained using a power-law size distribution with approximately the same index as is required to explain the interstellar reddening. The use of a narrower distribution leads to inconsistent grain sizes for these two sets of observations.

iii) The use of a broader size distribution function results in an increase in the polarising range, η (12.6), for all grain models. It then becomes possible to explain the variation of polarisation in NGC1999 and other nebulae, by a factor of 4 or more, using refractive indices of typical refractory materials. This is not possible with a narrower size distribution.

iv) A broad size distribution reduces the strength of the wavelength dependence of polarisation expected from a narrow size distribution to a value which is more consistent with typical observations. If a pure power-law is assumed, all variations must be accounted for by optical depth effects. Typical variations are consistent with the levels of depolarisation expected for small optical depths such as might be found in reflection nebulae (e.g. NGC1999, Chapter 14).

v) The scattering cross section for grains with a size distribution like equation 15.6 has the wavelength variations of equation 15.7. For scattering close to the illuminating star (where internal reddening between

the star and scattering centres is low) the expected nebular (B-V) colour difference (nebular colour - star colour) is -0.31 mag. using γ = 4.3 (the value which gives R = 3.3). This result is in good agreement with NGC1999 (-0.35 $\pm$ 0.2) and is typical of many nebulae.

vi) With a power-law distribution, it is the wavelength of observation which determines the apparent size of the dust grains which we observe. This thus provides an elegant explanation of the coincidence which results in the galaxy apparently being populated by grains with a size comparable to the wavelength of visible light and thus lying in the narrow range between Rayleigh scattering (constant high polarisations) and geometric scattering (low polarisations). A power-law has no "characteristic grain size" except, perhaps a cut off at small sizes which is not significant in the visual waveband.

vii) The observed constancy of grain properties throughout the galaxy is difficult to explain using a narrow size distribution, since the characteristic size must remain constant to within narrow limits. A power-law distribution, however, gives an apparent size which is determined by the wavelength of observation and is thus relatively insensitive to the effects of the grain environment (e.g. the effects of collisions in reducing grain sizes, or of accretion in increasing them are small). If, in the ultraviolet, the effects of a size cut-off

are important, far more variability would be expected, however.

Mathis et Al. have discussed the observations of interstellar polarisation in the context of their models. The size distributions resulting from their fits to the extinction data gave maximum interstellar polarisation at ultraviolet wavelengths. Clearly, however, the observation of maximum polarisation at visual wavelength indicates the presence of a "characteristic size" of grain in the interstellar medium. This may result, as Mathis et Al. suggest, because of a highly polarising component, of a characteristic size, which does not contribute significantly to the extinction, or it may be possible to introduce deviations from ideal power law distributions which have the characteristic size required while not significantly affecting the scattering properties. A third explanation could be that the particle shape is size dependent.

If the suggestion of Mathis et Al., that the size distribution of interstellar dust approximates to a power-law, is accepted, it becomes necessary to abandon the formation/destruction equation (Equation 11.16), which was originally formulated for ice grains, since it results in size distributions with a quasi-exponential cut-off which appears to be too strong. An alternative way of forming the size distribution must then be sought. Mathis et Al. suggest that a power-law may represent some other form of "most probable" or "equilibrium" condition.

There appears to be some evidence, however, that very typical grain properties are seen in regions where the dust has recently been formed:

Firstly, the work of Carty et Al. (1979) in interpreting polarimetric observations of Eta Carinae (Warren-Smith et Al., 1979) shows that a narrow size distribution appropriate to the interstellar extinction law is inconsistent with the observations and that a smaller mean size and broader size distribution is required. Visvanathan (1967) also concludes that there is very little wavelength dependence of polarisation in this object, again indicating a broad, quasi-power-law distribution. Secondly, observations of the cometary nebula NGC6729 (Warren-Smith and Scarrott, unpublished - see table 14) show polarisations of the same level as in the nearby nebula NGC6726/7 (Figure 9.2) and Serkowski (1969) reports a weak wavelength variation of the polarisation of this object. Thirdly, observations of bipolar nebulae by Cohen and Kuhi (1977) and Cohen (1977) show striking wavelength independence in polarisation and, although interpreted by the authors in terms of Rayleigh scattering, seem to be consistent with a power-law distribution of particles (particularly in view of the asymmetry present which is of the same form as in Eta Carinae, but is not expected with Rayleigh scattering). Finally, collections of dust particles in the upper atmosphere and near-Earth space reported by Hemenway (1973) show a size distribution which is power-law, with an index slightly above 4 and a cut-off

at a small size of ~0.03μm.

This material may be matter accreted by the sun, or cometary debris. Hemenway shows that the dust collections are consistent with Zodiacal Light observations and favours an origin in solar sun-spots where the temperature is sufficiently low for their formation. He also suggests that similar grains emitted by other stars may be of astrophysical significance.

Thus, in 4 classes of object where there appears to be emission of solid matter from stars into the interstellar medium, the observations suggest that a very broad spectrum of grain sizes are involved which may approximate to a power-law distribution. It is tempting to suggest that a power-law may always arise as the result of grain condensation in stellar atmospheres and it would be interesting to know if a mechanism exists which favours such a size distribution. If it does, it may provide an elegant and consistent explanation of the properties of interstellar dust.

SUMMARY AND CONCLUSIONS FOR PART 2

New polarimetric observations of the nebula NGC1999 in the Orion dark cloud have been made at an effective wavelength of 0.49μm and show a high degree of polarisation, reaching 30 to 40% at 50" from the central star, V380 Orionis. These levels of polarisation are typical of those found in a number of other reflection nebulae, including NGC6726/7, measurements of which were used in part 1 of this thesis.

From the available data it is concluded that NGC1999 is a simple reflection nebula, illuminated by a single star contained within it, and that the observed levels of polarisation arise by light scattering from small dust grains. Dense obscuration seen close to the illuminating star appears to have no effect other than the attenuation of the nebular light, and so is probably foreground - its effect seems to be localised to a small region of the nebular surface. An asymmetry in the level of polarisation suggests that the nebula may be tilted with respect to the plane of the sky.

The nature of light scattering by size distributions of sub-micron dust particles has been investigated and the conclusion drawn that a number of commonly used size distribution functions give very similar scattering properties. The important parameter of a size distribution appears to be its functional form at large sizes.

Simple theoretical arguments based on a general model of a reflection nebula have been developed to determine which observed properties of nebulae relate most directly to the parameters important in nebular models. Three aspects of polarisation data appear important: the overall level of polarisation, which relates to the effective mean size of the scattering dust grains; the spatial structure of the polarisation, which relates to the nebular geometry and the overall range of polarisation (maximum:minimum) which relates to a parameter η (the polarising range of the dust grains), determined by how forward throwing the scattering functions of the grains are. Data on the variations in nebular brightness and colour are considerably less reliable indications of important nebular parameters, particularly where the optical depths within the nebula are appreciable, but the overall brightness can be used to estimate the dust density.

Applying the above ideas to the nebula NGC1999, a model has been developed, using a single scattering approximation, based on an illuminating star below the surface of a very large dust cloud of uniform density, whose front surface is tilted with respect to the plane of the sky. This model can explain the polarisation and surface brightness of the nebula if the dust grains have a high albedo, but the resulting dust densities indicate that the effects of multiple scattering of the light must

be taken into account.

A new technique to perform the line of sight integration in an optically thick nebula, using the Monte-Carlo method, has been produced and sufficiently high numerical efficiency has been achieved to enable the polarisation to be calculated in nebulae with general geometries.

Application of this method to the model of NGC1999 shows that, in the presence of multiple scattering, the nebular brightness is considerably underestimated by using the single scattering approximation, but that for optical depths $\ll 1$ the amount of depolarisation expected is small and changes little with position. Use of the single scattering approximation thus leads to an overestimate of the dust density, but little error in the estimation of the size and scattering properties of the dust grains or of the nebular geometry. Multiple scattering calculations have also been performed to calculate the nebular (B-V) colour for the model and the results are consistent with the observations of Bruck (1974) and differ from the predictions of the single scattering approximation. In the presence of multiple scattering, it appears that polarisation data retains a high degree of information about the nebular geometry and dust properties, while brightness and colour data rapidly become difficult to interpret and unreliable indicators of nebular parameters.

The available data on brightness, polarisation, colour and the extinction of the illuminating star in NGC1999

are consistent with it being a reflection nebula formed by a star embedded to a depth of between 0.1 and 0.15 pc. below the surface of a very large dust cloud of approximately uniform density. The required dust density in the cloud, on the basis of the 3 models considered, is shown in table 11. Due to uncertainties about the size distribution of the dust grains, however, it is difficult to relate these figures to estimates of the mass density of dust, although a figure of ~ 1 to 3×10^{-19} kgm^{-3} may be an approximate lower limit for a density of 10^{3} kgm^{-3} for the grain material. The tilt of the front face of the dust cloud with respect to the plane of the sky (away from the observer in P.A.$\sim 40^{\circ}$) is not well defined, but probably lies towards the lower end of the range 17° to 52° given by the models. Tilts of this magnitude can explain the observed asymmetry in the polarisation but introduce little asymmetry in the brightness or colour of the nebula. The geometric parameters of the nebula, derived on the basis of 3 refractive indices for the dust material and a grain size distribution $n(a) \propto \exp\left[-5(a/a_{o})^{3}\right]$, are shown in table 9. Geometries of this type may be sufficiently common and easily modelled to make them an important tool in the investigation of dust in reflection nebulae.

The mean grain size of the dust, appropriate to a given refractive index, was very well defined by the models and figures are given in table 9. Dust models selected to fit the polarisation data also gave very

similar extinction and scattering colours and it is not possible to discriminate between the three refractive indices considered on the basis of brightness or colour data. The polarisation data, which was not fitted equally well by all grain models, may provide a future basis for discriminating between refractive indices, but in this study, uncertainties about the size distribution of the dust grains mean that differences in the quality of fits to the polarisation data are probably not significant. No conclusions about the dust material can thus be inferred from this study, except that the grain albedo must be close to unity, and the three refractive indices considered: $m = 1.16$, appropriate to the bacterial particles of Wickramasinghe (1979b); $m = 1.33$, appropriate to ice and $m = 1.63 - 0.05i$, appropriate to silicate material, all appear to be admissible. Further information about the size distribution of the dust grains is required before the grain material can be reliably investigated.

The dust grain sizes required to explain the observations of NGC1999 are significantly smaller than those required to fit the interstellar extinction law in the visual wavelength region, and similar small sizes would also be required to explain many other nebular polarisation observations. This inconsistency may either be resolved by supposing that nebular dust is physically altered by the presence of an illuminating star and is thus not typical of the general interstellar material,

or dust models may be formulated which result in the grains appearing smaller when observed by large angle scatterings than when observed by extinction (which includes small angle scatterings). Observations which suggest that dust has very similar properties in many nebulae, together with problems associated with explaining the observed range of polarisation in nebulae, may favour the second alternative and the use of multi-component dust models, core-mantle grains or a very broad distribution of grain sizes may all be possible solutions to the problem. Further numerical work is required to investigate these possibilities.

The use of a quasi-power-law size distribution (probably in conjunction with several dust components), as derived by Mathis et Al. (1977) on the basis of the interstellar extinction law, appears an attractive solution, as it may be able to reconcile the extinction and nebular polarisation data and can also explain many other features of the interstellar dust. Such a distribution may arise in the dust condensation process in stellar atmospheres. There remain problems in interpreting the interstellar polarisations using this model, however. An observational test, based on how the apparent size of dust grains in nebulae changes with the wavelength of observation, appears possible and further polarimetric observations of NGC1999 and other nebulae, at a variety of wavelengths, would be of great interest in determining the nature of the size distribution of interstellar dust so that tests of grain materials could be reliably carried out.

CONCLUDING REMARKS

A glance at any review of work on interstellar dust shows that, over many years, observations of reflection nebulae have contributed little to our understanding of this matter when compared with the many conclusions drawn on the basis of interstellar extinction and polarisation data. It is hoped that this thesis has indicated that the extent of information available from nebular observations is potentially very large, if the problems of interpretation, which have hitherto limited their exploitation, can be overcome. The angular scattering properties of dust can, at least, impose useful constraints on grain models, while at most it may be possible to trace out the size distribution function of the dust grains and ultimately determine their refractive indices at a number of wavelengths - information which extinction data has so far failed to provide.

The importance of imaging polarimetry in this work must be stressed, since it is the spatial information available in this technique which allows the single most important unknown in the interpretation of nebular data - the geometry - to be investigated. A reliable technique for imaging polarimetry, together with numerical techniques for the interpretation of the observations (including the effects of finite optical depth) are clearly essential if this work is to succeed, but they represent only the start of the investigation and observations at other wavelengths - both complete polarisation maps and

spectropolarimetry of selected nebular regions - are urgently needed.

Further data for NGC1999 in B and R wavebands and for other nebulae, including Eta Carinae, at a number of wavelengths, have recently been obtained by members of the Astronomy group in Durham and are currently awaiting reduction. Interpretation of the results is likely to be of great interest in furthering our understanding of the dust in these nebulae.

REFERENCES

Allen, D.A., 1973. MNRAS, 161, 145.

Andriesse, C. D., Piersma, Th. R. & Witt, A. N., 1977. Astr. Astrophys., 54, 841.

Artamonov, B. P. & Efimov, Yu. S., 1978. Soviet Astron. A.J., 11, 607.

Axon, D. J., 1977. Ph.D. Thesis, U. of Durham.

Bergh, S. van den., 1966. Astron. J., 71, 990.

Billingsley, 1975. In Picture Processing and Digital Filtering, Topics in Applied Physics, Vol. 6, Ed. Huang, T.S., Pub. Springer - Verlag.

Bingham, R. G., McNullan, D., Pallister, W.S., White, C., Axon, D. J. & Scarrott, S. M., 1976. Nature, 259, 463.

Breger, M., 1974. Ap. J., 188, 53.

Brück, M. T., 1974. MNRAS, 166, 123.

Carty, T. S., Perkins, H. G., Warren-Smith, R. F. & Scarrott, S. M., 1979. MNRAS, in press.

Coffeen, D. L. & Hansen, J. E., 1974. In Planets, Stars and Nebulae Studied with Photopolarimetry, Ed. Gehrels, T., Pub. U. of Arizona Press.

Cohen, M., 1973. MNRAS, 161, 105.

Cohen, M., 1977. Ap. J., 215, 533.

Cohen, M. & Kuhi, L. V., 1977. Ap. J., 213, 79.

Davidson, K. & Ruiz, M. T., 1975. Ap. J., 202, 421.

Elvins, A. & Hall, J. S., 1966. Lowell Obs. Bull., 6 (No. 135), 257.

Feynman, R. P., Leighton, R. B. & Sands, M., 1965. In The Feynman Lectures on Physics, Vol. III, Pub. Addison-Wesley.

Fitzgerald, M. P., Stephens, T. C. & Witt, A. N., 1976. Ap. J., 208, 709.

Garrison, L. M. & Anderson, C. M., 1977. Ap. J., 218, 438.

1978a. Ap. J., 221, 601.

1978b. Ap. J., 224, 535.

Gaustad, J. E., 1963. Ap. J., 138, 1050.

Gillett, F. C. & Stein, W. A., 1971. Ap. J., 164, 77.

Greenberg, J. M., 1966. In Spectral Classification and Multicolour Photometry, Proc. of IAU. Symp. No. 24, Ed. Loden, K., Loden, L. O. & Sinnerstad, U.

Greenberg, J. M. & Hanner, M. S., 1970. Ap. J., 161, 947.

Greenberg, J. M. & Hong, S. S., 1973. In The Dusty Universe, Ed. Field, G. B. & Cameron, A. G. W., Pub. Smithsonian Astrophys. Obs.

Greenberg, J. M. & Roark, T. P., 1967. Ap. J., 147, 917.

Hammersley, J. M. & Handscomb, D. C., 1975. Monte Carlo Methods, Pub. Methuen & Co., London.

Hanner, M. S., 1971. Ap. J., 164, 425.

Hanner, M. S. & Greenberg, J. M., 1970. Ap. J., 161, 961.

Hemenway, C. L., 1973. In The Dusty Universe, Ed. Field, G. B. & Cameron, A. G. W., Pub. Smithsonian Astrophys. Obs.

Herbig, G. H., 1960. Ap. J. Suppl., 4, 337.

Hong, S. S. & Greenberg, J. M., 1978. Astron. Astrophys., 70, 695.

Hoyle, F. & Wickramasinghe, N. C., 1977. Nature, 266, 241.

1978. Nature, 270, 323.

1979a. Biochemical Chromophor and the Interstellar Extinction at Ultraviolet Wavelengths, Preprint.

1979b. Paper presented at Royal Astronomical Society Meeting, U. of Durham, 1979.

Hulst, H. C. van de, 1957. Light Scattering by Small Particles, Pub. John Wiley & Sons.

Jonckheere, R., 1917. Mem. RAS, 61.

Jura, M., 1975. Astron. J., 80, 227.

Khalesse, B., Pallister, W. S., Warren-Smith, R. F. & Scarrott, S. M., 1979. MNRAS, in press.

Kuhi, L. V., 1974. Astron. Astrophys. Suppl., 15, 47.

Lynn, P. A., 1973. An Introduction to the Analysis and Processing of Signals, Pub. Macmillan, London.

Mathis, J. S., Rumpl, W. & Nordsieck, K. H., 1977. Ap. J., 217, 425.

Mattila, K., 1970. Astron. Astrophys., 9, 53.

McGee, J. D. & McMullan, D., 1969. Advances in Electronics and Electron Physics, 28A, 61.

McMullan, D., Powell, J. R. & Curtis, N. A., 1972. Advances in Electronics and Electron Physics, 33A, 37.

Mendoza, E. E., 1966. Ap. J., 143, 1010.

1968. Ap. J., 151, 977.

Menon, T. K., 1958. Proc. Inst. Radio Engrs., 46, 230.

Mie, G., 1908. Ann. Physik., 25, 377.

O'Dell, C. R., 1965. Ap. J., 142, 604.

Ohman, Y., 1939. MNRAS, 99, 624.

Oort, J. H. & Hulst, H. C. van de, 1946. Bull. Astron. Inst. Neth., 10, 187.

Pallister, W. S., 1976. Ph.D. Thesis, U. of Durham.

Patterson, T. N. L., 1968a. Math. Comp., 22, 847.

1968b. Math. Comp., 22, 877.

Penston, M. V., 1969. MNRAS, 144, 159.

Racine, R., 1968. Astron. J., 73, 233.

Reitz, J. R. & Milford, F. J., 1973. Foundations of Electromagnetic Theory, 2nd. Edition, Pub. Addison-Wesley.

Roark, T., Roark, B. & Collins, G. W. II, 1974. Ap. J. 190, 67.

Rogers, C. & Martin, P. G., 1979. Ap. J., 228, 450.

Rush, W. F., 1975. Astron. J., 80, 37.

Sakata, A., Nakagawa, N., Iguchi, T., Isobe, S., Morimoto, N., Hoyle, F. & Wickramasinghe, N. C., 1977. Nature, 266, 241.

Serkowski, K., 1969. Ap. J., 158, 1107.

Sobolev, V. V., 1960. Soviet Astron. A. J., 4, 1.

Strom, S. E., Strom, K. M. & Grasdalen, G. L., 1975. Ann. Rev. Astron. Astrophys., 13, 187.

Vanysek, V., 1969. Vistas in Astronomy, 11, 189.

Vanysek, V. & Solc, M., 1973. In Interstellar Dust and Related Topics, Proc. of IAU. Symp. No. 52, Ed. Greenberg, J. M. & Hulst, H. C. van de.

Visvanathan, N., 1967. MNRAS, 135, 275.

Voshchinnikov, N. V., 1977. Soviet Astron. A. J., 21, No. 6.

1978. Soviet Astron. A. J., 22, 188

Warren-Smith, R. F., Scarrott, S. M., Murdin, P. & Bingham, R. G., 1979. MNRAS, 187, 761.

Werner, M. W. & Salpeter, E. E., 1969. MNRAS, 145, 249.

White, R. L., 1979a. Ap. J., 229, 954.

1979b. Ap. J., 230, 116.

Wickramasinghe, N. C., 1973. Light Scattering Functions for Small Particles with Applications in Astronomy, Pub. Hilger, London.

Wickramasinghe, N. C., Hoyle, F. & Nandy, K., 1977. Astrophys. Space Sci., 47, L9.

Wickramasinghe, N. C. & Nandy, K., 1972. Rep. Prog. Phys., 35, 159.

Witt, A. N., 1977a. Ap. J. Suppl., 35, 1.

1977b. Ap. J. Suppl., 35, 7.

1977c. Ap. J. Suppl., 35, 21.

1977d. Ap. J. Suppl., 35, 31.

Witt, A. N. & Stephens, T. C., 1974. Astron. J., 79, 948.

Zellner, B. H., 1970. Astron. J., 75, 182.

Zellner, B., 1974. In Planets, Stars and Nebulae Studied with Photopolarimetry, Ed. Gehrels, T., Pub. U. of Arizona Press.

Zerull, R. & Giese, R. H., 1974. Ibid.

ACKNOWLEDGEMENTS

The author would like to thank his supervisor in Durham, Dr. S. M. Scarrott, for his support in carrying out the work described in this thesis, and in particular for the organisation and planning involved in the collaborative projects and observing visits abroad, on which Observational Astronomy depends, and for providing the stimulus for this work in Durham.

Particular thanks are also due to Dr. W. S. Pallister for his collaboration and unending enthusiasm in the development of the polarimeter reduction technique and for his assistance in checking and improving many computer programmes.

The author would also like to thank all those on whose assistance he has depended during this work - the other members of the Astronomy group in Durham, members of the Physics Department, all those associated with the running of the NUMAC computing service in Durham and the staff of the Royal Greenwich Observatory and of the observatories in Israel, Australia and South Africa which the author has visited. Particular thanks are extended to those whose hospitality has eased the burden of working away from home.

The University of Durham and the Science Research Council are acknowledged for their financial support both in Durham and during visits abroad.

Finally, thanks are due to Carol Jackson for her support and encouragement during the writing of this thesis and for checking the manuscript and to Mrs. A. Smiles for her speedy and accurate typing.